

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 120.24 Seconds  
(without alignments)  
1738.900 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIYVGDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	6	ABP73016 Amino aci
2	4036	100.0	740	7	ADD22926 Acidother
3	4036	100.0	740	7	ADD22923 Acidother
4	4036	100.0	957	6	ABP73015 Amino aci
5	4036	100.0	957	6	ADD22921 Acidother
6	2063	51.1	940	6	Abb99489 Amino aci
7	1680	41.6	726	7	ADD22927 Aspergill
8	1579	39.1	838	7	ADD42055 Trichoder
9	841.5	20.8	555	7	ADD24919 Xanthomon
10	238	5.9	2468	6	ABU38411 Protein e
11	238	5.9	2468	6	ABP59933 Microbia
12	199.5	4.9	2435	6	ABU19529 Protein e
13	181.5	4.5	3892	6	ADA34216 Acinetoba
14	181	4.5	1308	6	ABU16642 Protein e
15	170	4.2	1074	6	ABU22692 Protein e
16	168.5	4.2	1439	6	ABU16643 Protein e
17	168.5	4.2	5291	7	ADC01014 Enterohae
18	167.5	4.2	1119	7	ABW01170 Candida a
19	166	4.1	1468	7	ADD48744 Rat Prote
20	163.5	4.1	1289	6	ABU16977 Protein e
21	163	4.0	1684	2	AAR14948 Bacterial
22	160.5	4.0	3716	6	ABM15900 Mycobacte
23	159.5	4.0	1291	2	AAM59912 Amino aci
24	158.5	3.9	774	2	AAW35390 Flavobact
25	158.5	3.9	1440	5	ABB54801 Lactococc

ALIGNMENTS

RESULT 1

ABP73016  
ID ABP73016 standard; peptide; 740 AA.

XX AC ABP73016;

XX XX 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the avicelase AvIII catalytic domain.

XX XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;

KW KW detergent; pulp processing; paper processing; feed processing; textile;

XX KW cellulose.

XX OS Acidothermus cellulolyticus.

XX XX WO2003012090-A2.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023818.

XX PR 28-JUL-2001; 2001WO-US023818.

XX XX (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX DR WPI; 2003-248177/24.

PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful

PT for degradation of cellulose or in generating anti-AvIII antibodies for

PT purifying recombinant AvIII polypeptides from genetically engineered

XX host cells.

XX PS Claim 6; Page 8; 4pp; English.

XX CC The present sequence is derived from a thermostable avicelase, designated

CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,

CC and is a cellulase. AvIII is useful in the conversion of biomass to

CC biofuels and biofuel additives. It may be useful in the production of

CC detergents, pulp and paper processing, food and feed processing and in

CC textile processes. The thermostable AvIII peptide is useful in the

CC degradation of cellulose, and in generating specific anti-AvIII

CC antibodies that are useful in purifying recombinant AvIII polypeptides

CC from genetically engineered host cells, in detecting AvIII polypeptide

CC expression, as well as a reagent tool for characterizing the molecular

CC actions of the polypeptide. The AvIII polynucleotide is useful as a

ABU36440 Protein e  
ABP43908 MUC5B par  
ADE08753 Novel pro  
ABP65361 Bifidobac  
AAV13493 Truncated  
AAE16324 Active ce  
AAP90615 Polypepti  
ABU19388 Protein e  
AAE37932 Human CGD  
ABU34624 Protein e  
AAV90913 Cenoarchae  
AAY78516 Isoamylas  
ABU17029 Protein e  
ABP69842 Human pol  
ABP69841 Human pol  
ABP69840 Human pol  
AAU31850 Novel hum  
ABU34548 Protein e  
ABU39844 Protein e  
AAP94419 Primary t

CC source of probes or primers in various diagnostic assays

XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 6; Length 740;  
Best Local Similarity 100.0%; Pred. No. 8.4e-271; Indels 0; Gaps 0;  
Matches 740; Conservative 0; Mismatches 0;

Qy 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Qy 61 NNWYGVVSTAAADPINTNKWAAVGMVYTNWDNDGAILRSSDQATWQITPLPFKLG 120  
Db 61 NNWYGVVSTAAADPINTNKWAAVGMVYTNWDNDGAILRSSDQATWQITPLPFKLG 120  
Qy 121 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240  
Qy 241 IPHKGVPDPVNVHLYIATSNFGPYDGSNGVWKFVSTGVTWTRISPVSTDTANDYFGY 300  
Db 241 IPHKGVPDPVNVHLYIATSNFGPYDGSNGVWKFVSTGVTWTRISPVSTDTANDYFGY 300  
Qy 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360  
Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360  
Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Db 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Qy 421 APWKGLEETAANDLIIPPSPGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
Db 421 APWKGLEETAANDLIIPPSPGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
Qy 481 AELNPSIIVRAGSDPPSQPNDRHVAFTDCKWKFQSEFGVTTGTTVAASADGRFV 540  
Db 481 AELNPSIIVRAGSDPPSQPNDRHVAFTDCKWKFQSEFGVTTGTTVAASADGRFV 540  
Qy 541 WAPGDQPQVYVYVGFNSWAAASQGVANAOIRSDRVNPKTFYALNCTFYRSTDDGGVTF 600  
Db 541 WAPGDQPQVYVYVGFNSWAAASQGVANAOIRSDRVNPKTFYALNCTFYRSTDDGGVTF 600  
Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Qy 661 KSAPGSSYPVAVVGTGTVGAYRSDCCGTTWLVINDDQHQYGNWQAITGDHANLRRV 720  
Db 661 KSAPGSSYPVAVVGTGTVGAYRSDCCGTTWLVINDDQHQYGNWQAITGDHANLRRV 720  
Qy 721 YIGTNGRGIVYDGGAPSG 740  
Db 721 YIGTNGRGIVYDGGAPSG 740

RESULT 2

ADD22926

ID ADD22926 standard; protein; 740 AA.

XX AC

ADD22926;

XX DT

15-JAN-2004 (first entry)

XX DE

Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX KW

enzyme; AvIII; cellulose reduction; agricultural biomass;

XX MW

municipal solid waste; glycoside hydrolase; avicelase.

XX Acidothermus cellulolyticus.

XX US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX PA (DING/) DING S.

PA (ADNEY/) ADNEY W S.

PA (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

XX New isolated thermal avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX Example 2; SEQ ID NO 6; 29pp; English.  
XX The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
XX cellulolyticus avicelase AvIII catalytic domain.

XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 7; Length 740;

Best Local Similarity 100.0%; Pred. No. 8.4e-271; Indels 0; Gaps 0;

Matches 740; Conservative 0; Mismatches 0;

Qy 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Qy 61 NNWYGVVSTAAADPINTNKWAAVGMVYTNWDNDGAILRSSDQATWQITPLPFKLG 120  
Db 61 NNWYGVVSTAAADPINTNKWAAVGMVYTNWDNDGAILRSSDQATWQITPLPFKLG 120  
Qy 121 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240  
Qy 241 IPHKGVPDPVNVHLYIATSNFGPYDGSNGVWKFVSTGVTWTRISPVSTDTANDYFGY 300  
Db 241 IPHKGVPDPVNVHLYIATSNFGPYDGSNGVWKFVSTGVTWTRISPVSTDTANDYFGY 300  
Qy 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360  
Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWRIWDWTSYPNRSRYVLDISAE 360  
Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Db 361 PWLTFGVQPNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Qy 421 APWKGLEETAANDLIIPPSPGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480



Db 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQGSBPGVTTGTTVAASADGSRFV 540  
Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQGSBPGVTTGTTVAASADGSRFV 540  
Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 600  
Db 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 600  
Qy 601 QPVAAGLPSSGAVGVMFHAFVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGVMFHAFVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Qy 661 KSAFGSSYPVAVFVGTGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Db 661 KSAFGSSYPVAVFVGTGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Qy 721 YIGTNGRGIVYDGGAPSG 740  
Db 721 YIGTNGRGIVYDGGAPSG 740  
RESULT 3  
ADD22923  
ID ADD22923 standard; protein; 740 AA.  
XX AC ADD22923;  
XX DT 15-JAN-2004 (first entry)  
XX DE Acidothermus cellulolyticus avicelase Aviii catalytic domain.  
XX KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
XX KW municipal solid waste; glycoside hydrolase; avicelase.  
XX OS Acidothermus cellulolyticus.  
XX PN US2003108988-A1.  
XX PD 12-JUN-2003.  
XX PF 18-OCT-2002; 2002US-00155400.  
XX PR 28-JUL-2001; 2001US-00917376.  
XX PA (DING/) DING S.  
XX PA (ADNEY/) ADNEY W S.  
XX PA (VINZ/) VINZANT T B.  
XX PA (HIMM/) HIMMEL M E.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX PS WPI; 2003-810853/76.  
XX PT New isolated thermal tolerant avicelase polynucleotide useful for  
XX PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
XX PT in a starting material, e.g. municipal solid waste.  
XX PS Claim 16; SEQ ID NO 3; 29pp; English.  
XX CC The invention relates to an isolated polynucleotide molecule encoding a  
XX CC thermostable Aviii polypeptide. The polynucleotide is useful for  
XX CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
XX CC useful for reducing cellulose in a starting material which involves  
XX CC administering to the starting material, e.g. agricultural biomass or  
XX CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
XX CC method further comprises administering a second polypeptide molecule  
XX CC chosen from the glycoside hydrolase family of proteins. The present  
XX CC sequence represents the amino acid sequence of Acidothermus  
XX CC cellulolyticus avicelase Aviii catalytic domain.  
XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 7; Length 740;  
Best Local Similarity 100.0%; Pred. No. 8.4e-271;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATTOPYTWSNVAIGGGFVDGIVFNEGARGILYVETDTCGMWRDAANGRWITPLLDWVGW 60  
Db 1 ATTOPYTWSNVAIGGGFVDGIVFNEGARGILYVETDTCGMWRDAANGRWITPLLDWVGW 60  
Qy 61 NNWGVNGVSTAADPINTNKVWAAVGMTNSWDPNDGAILRSSDQATWQITPLPKLGG 120  
Db 61 NNWGVNGVSTAADPINTNKVWAAVGMTNSWDPNDGAILRSSDQATWQITPLPKLGG 120  
Qy 121 NMPGRGMRGLAVDPNNDNIIYFGAPSKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGMRGLAVDPNNDNIIYFGAPSKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Qy 181 TGYQSDIQGVVWVAFDSSSSILGQASKTIFGVADPNPNFVWSRDGGATWQAVPGAPTGF 240  
Db 181 TGYQSDIQGVVWVAFDSSSSILGQASKTIFGVADPNPNFVWSRDGGATWQAVPGAPTGF 240  
Qy 241 IPHKGVEDPVNHLVYIATNTGGPYDGGSDGVKFSVTSGTWRISPVSTDTANDYFGY 300  
Db 241 IPHKGVEDPVNHLVYIATNTGGPYDGGSDGVKFSVTSGTWRISPVSTDTANDYFGY 300  
Qy 301 SGLTIDROHPNTIMVATOISWMPDTIIIFRSTDGGATWTRINDWTSYPNRSRLRYLDISAE 360  
Db 301 SGLTIDROHPNTIMVATOISWMPDTIIIFRSTDGGATWTRINDWTSYPNRSRLRYLDISAE 360  
Qy 361 PWLTFGVQPNPPVPSKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 420  
Db 361 PWLTFGVQPNPPVPSKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 420  
Qy 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
Db 421 APMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 480  
Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQGSBPGVTTGTTVAASADGSRFV 540  
Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQGSBPGVTTGTTVAASADGSRFV 540  
Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 600  
Db 541 WAPGDPQPVVYAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 600  
Qy 601 QPVAAGLPSSGAVGVMFHAFVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGVMFHAFVKGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Qy 661 KSAFGSSYPVAVFVGTGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Db 661 KSAFGSSYPVAVFVGTGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Qy 721 YIGTNGRGIVYDGGAPSG 740  
Db 721 YIGTNGRGIVYDGGAPSG 740  
RESULT 4  
ABP73015  
ID ABP73015 standard; protein; 957 AA.  
XX AC ABP73015;  
XX DT 03-JUN-2003 (first entry)  
XX DE Amino acid sequence of the avicelase Aviii.  
XX KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX KW detergent; pulp processing; paper processing; feed processing; textile;  
XX KW cellulose.  
XX OS Acidothermus cellulolyticus.



CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase AvIII.

SQ Sequence 957 AA;

Query Match 100.0%; Score 4036; DB 7; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1.2e-270;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPTWSNVAICGGFVDCGIVNEGAPGLYVRTDIGMYRWDAAANGRIWIPLLDWGV 60  
DB |||||  
QY 47 ATTQPTWSNVAICGGFVDCGIVNEGAPGLYVRTDIGMYRWDAAANGRIWIPLLDWGV 106  
DB |||||  
QY 61 NNWYNGVVSIAADPINTNKVAAVGYMYSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
DB |||||  
QY 107 NNWYNGVVSIAADPINTNKVAAVGYMYSWDPNDGAILRSSDQATWQITPLPFLKGG 166  
DB |||||  
QY 121 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYIANPTDT 180  
DB |||||  
QY 167 NMPGRGGERLAVDPNDNDNIIYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYIANPTDT 226  
DB |||||  
QY 181 TGYQSDIQGVVWAFDKSSSLGQASKTI FGVADPNNPVFWSRGGATWQAVCAPTGF 240  
DB |||||  
QY 227 TGYQSDIQGVVWAFDKSSSLGQASKTI FGVADPNNPVFWSRGGATWQAVCAPTGF 286  
DB |||||  
QY 241 IPHKGVPDPVNNHLYIATSTGGPYDSSGVDWKFVSTGVTWTRISPVPSDTDTANDYFGY 300  
DB |||||  
QY 287 IPHKGVPDPVNNHLYIATSTGGPYDSSGVDWKFVSTGVTWTRISPVPSDTDTANDYFGY 346  
DB |||||  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIPRSTDDGATWRIWDTSYNRSRLRYVLDISAE 360  
DB |||||  
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIPRSTDDGATWRIWDTSYNRSRLRYVLDISAE 406  
DB |||||  
QY 361 PWTFTGVQPNPPVPSKLGWDEAMADPNSDRMLYGTGATLYATNDLTWKWDSGGQTHI 420  
DB |||||  
QY 407 PWTFTGVQPNPPVPSKLGWDEAMADPNSDRMLYGTGATLYATNDLTWKWDSGGQTHI 466  
DB |||||  
QY 421 APWKVGLTEAVNDLIIPSPGAPLISALDGLGGFTHADVTAVPSTIFTSPTFTGTSVDY 480  
DB |||||  
QY 467 APWKVGLTEAVNDLIIPSPGAPLISALDGLGGFTHADVTAVPSTIFTSPTFTGTSVDY 526  
DB |||||  
QY 481 AELNPSIIVRAGSFPSSQPNDRHVAFSTGGKWFQSEPCGVTTGTTVAASADGSRFV 540  
DB |||||  
QY 527 AELNPSIIVRAGSFPSSQPNDRHVAFSTGGKWFQSEPCGVTTGTTVAASADGSRFV 586  
DB |||||  
QY 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFRSTDDGGVTF 600  
DB |||||  
QY 587 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFRSTDDGGVTF 646  
DB |||||  
QY 601 QPVAAGLPSSGAVGMFHAVPKGGLWLAASGLYHSTNGSSWSALTGVSSAVNVVGF 660  
DB |||||  
QY 647 QPVAAGLPSSGAVGMFHAVPKGGLWLAASGLYHSTNGSSWSALTGVSSAVNVVGF 706  
DB |||||  
QY 661 KSAFGSSYPVAVVGTIGVGTGAYRSDCGTWTWVLINDDQHQYGNWGAITGDHANLRRV 720  
DB |||||  
QY 707 KSAFGSSYPVAVVGTIGVGTGAYRSDCGTWTWVLINDDQHQYGNWGAITGDHANLRRV 766  
DB |||||  
QY 721 YIGTNGRGIVYDITGGAPSG 740  
DB |||||  
QY 767 YIGTNGRGIVYDITGGAPSG 786  
DB |||||

RESULT 6

ABB99489

ID ABB99489 standard; protein; 940 AA.

XX

AC

ABB99489;

XX

DT 03-MAR-2003 (first entry)  
XX Amino acid sequence of a xyloglucanase enzyme.  
DE Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;  
KW textile scouring; cellulose fiber processing; retting.  
XX Jonesia sp.  
OS  
XX WO200277242-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-DK000210.  
XX  
PR 27-MAR-2001; 2001DK-00000504.  
XX  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Duffner F, Sjöholm C;  
XX  
DR WPI; 2003-092855/08.  
DR N-PSDB; ABV76941.  
XX  
PS New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,  
PT and endogenous to a bacterium, useful in the textile industry for  
PT improving properties of cellulosic fibers, yarn or fabric.  
XX  
PS Claim 5; Page 72-75; 76pp; English.

CC The present sequence represents a xyloglucanase enzyme, belonging to  
CC family 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.  
CC DSM14140. The enzyme is useful in processes for machine treatment of  
CC fabrics. It is also useful in the textile industry for improving the  
CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in  
CC a textile scouring process step. The xyloglucanase enzyme is also useful  
CC in the cellulose fiber processing industry for retting of fibers such as  
CC hemp, jute, flax and linen. It is useful for preventing binding of  
CC certain soils to the xyloglucan left on the cellulosic material

SQ Sequence 940 AA;

Query Match 51.1%; Score 2063; DB 6; Length 940;  
Best Local Similarity 52.1%; Pred. No. 5.2e-134;  
Matches 386; Conservative 112; Mismatches 225; Indels 18; Gaps 9;

QY 6 YTWNSVAICGGFVDCGIVNEGAPGLYVRTDIGMYRWDAAANGRIWIPLLDWGVNNGY 65  
DB |||||  
QY 44 YSWSNVEIVGGFVDCGIVNEGAPGLYVRTDIGMYRWDAAANGRIWIPLLDWGVNNGY 103  
DB |||||  
QY 66 NGVVSIAADPINTNKVAAVGYMYSWDPNDGAILRSSDQATWQITPLPFLKGGNMPGR 125  
DB |||||  
QY 104 SGILSLATDPVDTNRVYLAAGTYSMDWDPQNGAILRSSDQATWQITPLPFLKGGNMPGR 163  
DB |||||  
QY 126 GNGERLAVDPNDNDNIIYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYIANPTDTTG-YQ 184  
DB |||||  
QY 164 GNGERLAVDPNDNDNIIYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYIANPTDTTG-YQ 220  
DB |||||  
QY 185 SDIQGVVWAFDKSSSLGQASKTI FGVADPNNPVFWSRGGATWQAVCAPTGFPHK 244  
DB |||||  
QY 221 GQNGQVWVTFDPTSAKAGQTTOTIIVGVADKQNNVYRSTDDGATWQAVCAPTGFPHK 280  
DB |||||  
QY 245 GVFPDPVNHVLYIATSTGGPYDSSGVDWKFVSTGVTWTRISPVPSDTDTANDYFGYSLT 304  
DB |||||  
QY 281 GVLDDHKGQQLYIATSDTGGPYDSSGVDWKFVSTGVTWTRISPVPSDTDTANDYFGYSLT 339  
DB |||||  
QY 305 IDROHPNTIMVATQISWMPDTIIPRSTDDGATWRIWDTSYNRSRLRYVLDISAEPLT 364  
DB |||||  
QY 340 IDRKNPDTIMVVSQVSWPDMYVIRSTDRGKTWSPWELNGSQPRTKQYNHDSGAPWLD 399  
DB |||||  
QY 365 FGVQPNPPVPSKLGWDEAMADPNSDRMLYGTGATLYATNDLTWKWDSGGQTHIAPMV 424  
DB |||||  
QY 400 FGNTAKEPEANPKLGWMTQSFIDPHNSDRFFYGTGAGIYGGTNTLNWWDKGGKVDITVKA 459  
DB |||||

[illegible]

RESULT 7  
ADD22927  
ID ADD22927 standard; protein; 726 AA.

AC ADD22927;

15-JAN-2004 (first entry)

DE *Aspercillus aculeatus* avicelase III catalytic domain.

XX  
KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
KW municipal solid waste: glycoside hydrolase: avicelase.

XX OS Aspergillus aculeatus.

XX  
PN  
US2003108988-A1.

AA  
PD  
12-JUN-2003.

18-OCT-2002: 2002US-00155400.

XX 28-JUL-2001: 2001US-00917376.

AA  
PA  
(DING/) DING S.

PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B

XX  
PA (HIMM/ ) HIMMEL M E.

PL 111-118, ADRIAN W.S.,  
XX

DR  
XX  
WPI; 2003-810853/16.

PT New isolated thermal  
PT detection of a polyn

PI in a starting material, e.g. municipal solid waste.

PS Example 2; SEQ ID NO 7; 29pp; English.

XX The invention relates to an isolated polynucleotide

CC

CC detection of a polynucleotide encoding a thermostable AVIII polypeptide. The

usertl for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents the amino acid sequence of Aspergillus aculeatus

CC avicelase III catalytic domain.

Sequence 726 AA;

Query Match 41.6%; Score 1680; DB 7; Length 726;  
Best Local Similarity 46.2%; Pred. No. 1.2e-107;  
Matches 346; Conservative 113; Mismatches 258; Indels 3

Qy	1	ATTOPYTWNSVAI--GGCGFFVDGIYVNPBCAGILYVRTDIGMYRWDAANGRWIPLLDWVG	59
Db	1	AASQAYTKNVTGGGGGFTPGIYVNFPSAKGVAYARTDIGAYRLN--SDDTWTPMLMDWVG	59
Qy	60	--WNWNGYNGWSIAADPINTNKVMAAVGMVWYSWDPNDGAILRSSDQAGTQWLTPLPF	116
Db	60	NDTWHDW--GIDALATDPVDDTRVYVAVGMVWYNWDPNVCGLIRSTDQGTWTETKLPF	116
Qy	117	KLGNMPPRGMBERLAVDPNNNDILYFCAPSCKGLMRSTDGATWSQMTNFPDVGTYIAN	176
Db	117	KVCGNMPRGMBERLAVDPNKNLSILYFGARSGHGLWKSTDYATGWSNVYTSPTWTGTYFD	176
Qy	177	PTDITGQSDIQGVVWYAFPKSSSSSLCQASKYIIFGVADPNPNVPWSDGGATWQAVPCA	236
Db	177	SSST--YTSDPVGIAWYFTDSTGSSGSAATPRIIFGVADAGKSPFKSBDAGATWAWSGE	234
Qy	237	PT-GFIHPKGVDFDPNVHLVIATISNTGCPYDSSGDVWKPVSTGWTNRISVPVSTDTAN	295
Db	235	POYGFLPHKGVLSPEEKLIIYSYANGAGPYDGTNGTVHKYNIITSGVWTDISP--TSLAS	291
Qy	296	DYFGYSGLITDRQHPNTIMVATQISWMPDITIIPRSTDGGATWTRIDWTSYPNRSRLYYL	355
Db	292	TYYGCGGLSVDLQVPGTLMAALNCWMPDELIIPRSTDGATWSPRIEWNGYPSINYYSY	351
Qy	356	DISAEPWLITGVQPNP--PVSPKLGWMDMAAIDPFNSDRDMLYGTGATLYATNDLTWDS	414
Db	352	DISNAPIQDITSTTDQFPV--RVGMVVEALAIIDPFDSNHLWLYGTGLTVYVGHDITNWS	408
Qy	415	GGQIHIAPMVKGLBETAUNDLISPPSCAPLISALGDLGGFTHADVTPSITITFTSPVFT	474
Db	409	KHNVTVKSLAVGIEEMAVLGLITPPGGPALLUSVAGDDGGFVHSDLDAAPNQAHYHTPYGT	468
Qy	475	GTSVDYAEALNPSIIVRAGSDPPSQPNDRHVAFSTDGCKWFOGSEPGGVTTGGTVAAASA	534
Db	469	TNGIDYAGKPSNIVRSGSD--DYPT--LALSSNFGSTWYADYAASTSTCTGAVALSA	523
Qy	535	DGSEFWAPGDGPQPVYVAVGFNGSWAASQGVPAQAIRSDRVNPKTPYALNNGTFYRST	594
Db	524	DGDTVLMLMSSTGALVSKSQG--TLTAVSSLPSGCAVIASDKSDNTVPYGGSAGAIYVSK	580
Qy	595	DGGVTTFQVAAGLPSSGAVGMFHAVPKGDLWLAASSGLVHSTNGGSSNSAI--TGVS	653
Db	581	NTATSFTKTVS--JGSSTTVNAI--RAHPSIAGDVWASTDKGLWHSHTDYDGTFTQIGSGVTA	638
Qy	654	AVNVGFGKSPAGSSYPVAVFVGTIGTVGAYRSDDCGTTWVLINDDQHYQWNGQA--IT	711
Db	639	GWSFGFGKASTGCVYVIYGFFTIDGAAGLFXKSEDAGTNNQVSDASHGFGS--GSANVN	697
Qy	712	GDHANLRVYIGNRGIVYDGIIGAPSG	740
Db	698	GDLOTYQGVFRGHPGHLROSOREPAG	726

## RESULT 8

RESOUR 8  
ADD42055

ID ADD42055 standard; protein; 838 AA.

AC ADD42055;

DT 15-JAN-2004 (first entry)

DE Trichoderma reesei Family 74 xyloglucanase.

Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic link  
xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;

ethanol production; detergent composition; fabric treatment;  
textile treatment; enzyme.

Hypocrea jecorina.  
WO2003089598-A2.  
30-OCT-2003.  
17-APR-2003; 2003WO-US011831.  
19-APR-2002; 2002US-0373987P.  
(NOVO) NOVOZYMES BIOTECH INC.  
Michael R. Zaretsky E, Haas J;  
WPI; 2003-845528/78.  
N-PSDB; ADD42054, ADD42060.  
New polypeptides having Family 74 xyloglucanase activity, and encoding  
nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
-containing biomass to ethanol or as a detergent.  
Claim 8; SEQ ID NO 2; 96pp; English.

The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
fungus *Trichoderma reesei* (Hypocrea jecorina), and nucleic acids encoding  
it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
The invention also relates polypeptide sequences at least 70% identical  
to the enzyme, expression vectors and host cells comprising a nucleic  
acid of the invention, the recombinant production of the enzyme, and  
mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
the invention can be used in the degradation of cellulose- and  
hemicellulose-containing biomass to produce ethanol. It can also be used  
in a detergent composition for treating fabric during a machine washing  
cycle. The present sequence represents the *Trichoderma reesei* Family 74  
xyloglucanase.

XX SQ Sequence 838 AA;

Query Match 39.1%; Score 1579; DB 7; Length 838;  
Best Local Similarity 43.0%; Pred. No. 1.5e-100;  
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;

6 YTSNVAI-GGGFVGDIVFNEGAPGILYVTDIGMYRMDAANGRIPLDWW-...GW 60  
21 FSWKNVKGGGGFPVPGIIFHPKTKGVAYARTDIGLYRLN-ADDSWATVTDGIADNAGW 79  
61 NNWYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLGG 120  
80 HNW---GIDVALDPQDDQKYAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLGG 136  
121 NMPGRGGERLAVDPNNDNILYFCAPSGKGLWRGTDGATWSOMTNPDPVGTYIANPTDT 180  
137 NMPGRGGERLAVDPANSLIYFGARGNGLWKTDGCVTFKSVSTATGTATYIPDFSDS 196  
181 TGVQSDIQGVVWVAFDSSSLGQASKITIFGVADP-NNPVFWRDGGATQAVPGAPTG 239  
197 NGYNSDKQGLMVTFTDSTSTTGCATSRIFVGTADNITASVYVSTNAGSTWSAVFGQPGK 256  
240 FIPKGVDPVNHVLIYATNTGPGYDGSQDVWKFVSTGWTWRIKSPVSTDTANDYFG 299  
257 YFPHKAKLQPAEKALYLYTSDGTGPGYDGLGWSVRYDIAGTGWKIDITPVSGSDL-...YFG 313  
300 YSLGTIDROHPNTIMVATQISWPDPTTIFRSTDCGATWTRIDWTSYPNRSLRYVLDISA 359  
314 FGLGLDLQKPGTLVVASLNSWNPDAQFRSTDSGTWSPINAWASTPTETYYISITPK 373  
360 EPMWTFG-VQNPFPVPS-...PKLGMWDEAMAIIDFPNSDRMLYGTGATLYATNDLTKWDS 414  
374 APWKNNFIDVTSPPSGLIKRLIGLMWIESLEIDPTDSNHWLYGTGWTIFGHDLTWDT 433

QY 415 GQIHIAPMVKGLERTAVNDLISPPSGAFLISALGLDGGFTHA---DVTAVPSTIFTSPV 471  
DB 434 RHNVSIQSLADGIEEFVSQDLASA PGSELLAAVGDNDGFTFASRNDLGTSPQVWATPT 493  
QY 472 FTTGTSVDYAEINPSIIVRAGSDFDSSQPNDRHVARFSTGGKWFQGSSEPGVTTGGTVA 531  
DB 494 WATSTSVDYAGNSVSVVRVGNVTAGTQQ-----VAISSDGGATWSIDYAADTSMNGGTVA 548  
QY 532 ASADGSRFVWAPDGPQPVVYAVGVGNSWAASQGVPAQAIRSDRVNPKTFYALSNCTFY 591  
DB 549 YSADGDTILWSTASSG---VQRSQFGSFPASVSSLPAGAVIAKKTNSVFIAGSGSTFY 605  
QY 592 RSTDCGVTFQPVAAGLPSSGAVGVN---FHAVPGKEGDLMLAASSGLYHSTNGSSWSAI- 648  
DB 606 VSKDTGSSP---TRG-PKLSAGATTIRIDIAAHPPTAGTLYVSTDVGFIRSTDSGTTFQGS 661  
QY 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVTGA---YRSDDCGTTWVLINDHQHGVNW 706  
DB 662 TALNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQFGSID 717  
QY 707 GOAITGDHANLRRVYVGTNGRGIVY--GDIGGAPSG 740  
DB 718 STKVAGSGTAGQVYVGTNGRGVFAQGTGVTGGTGG 753

RESULT 9

ADD24919  
ID ADD24919 standard; protein; 555 AA.  
XX AC ADD24919;  
XX DT 15-JAN-2004 (first entry)  
XX DE Xanthomonas campestris cellulase #1.  
XX KW Directed genetic engineering; galactomannanase; reduced activity;  
KW enhanced activity; xanthan gum production; suspension stability;  
KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
KW cellulase; extracellular protease; intracellular protease;  
KW glucose dehydrogenase; enzyme.  
XX OS Xanthomonas campestris.  
XX FN US2003036176-A1.  
XX PD 20-FEB-2003.  
XX PF 10-AUG-2001; 2001US-00927827.  
XX PR 28-MAR-2001; 2001US-0279493P.  
XX PA (BOWE/) BOWER S G.  
XX PA (RAMS/) RAMSEIER T M.  
XX PI Bower SG, Ramseier TM;  
XX XX WPI; 2003-625389/59.  
DR N-PSDB; ADD24893.  
PT New transformed cell or organism having reduced or enhanced activity of  
PT at least one protein, useful for producing xanthan gum, which are useful  
PT for providing formulations and properties.  
XX PS Claim 1; SEQ ID NO 47; 135pp; English.  
XX CC The present invention relates to polypeptide and polynucleotide sequences  
CC from Xanthomonas campestris which may be used for activity reduction or  
CC enhancement using directed genetic engineering. A transformed cell or  
CC organism having reduced or enhanced activity of at least one such protein  
CC e.g. galactomannanase can be generated by disrupting the gene encoding  
CC the protein. The activity of the protein is reduced by the presence of an  
CC antisense nucleic acid sequence. The nucleic acid sequence of the gene



Db 1275 PIGETADSGNWSFTPGTFLANGTVVNAVAQDPAGNTPQGSTTVDAVAPNTPVNPSPN 1334  
 Qy 252 HVLVIATNTGGPY---DGSSGDVWKFVSVT-SGTW--TRISPVPS-----TDTANDYFGYS 301  
 Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIQQTADSGNWSFTPGSQLPNGTVVNVVTSADAAGNT 1394  
 Qy 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDDG-----ATWTRIDW 342  
 Db 1395 SLPATTVDSSLSPIQVDPNSGVSIGTADAGNTIII--TDGNGNPIQVTVADSGNWS 1452  
 Qy 343 WT-----SYNRSRLRYVLDISAEPLW--TFGVQPNPPVPSPKLGWDEMAIDPFNS 392  
 Db 1453 FTPGIPLPDGTVVNVVARSNSVDSAPAVITVDGAPAPV-----IDPSNG 1499  
 Qy 393 DRMLYGT---GATLYATNDLTWKDSGGQIHIAPMVKGLEETAVNDLISP---PSGAPLI 445  
 Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIQATADSGNWTFTPTPLANGTVIN 1548  
 Qy 446 SALGDLGGFTH-----ADVTAVPSTIF--TSPVFTTGTSDVVAELNPSIIVRAGSPDPS 497  
 Db 1549 AVAQDPAGNTSGPASVTVDAIAPPAPVNPNSGVVISGT----AEAGATVILTIDGNGNPI 1604  
 Qy 498 SQPNDRHFAFSTDGKNWF-----QSEPPGGVT----- 525  
 Db 1605 GQ-----VTADSGNWSFTPGTFLANGSVINALAQDAAGNNSPTSATVDSLAPAPV 1657  
 Qy 526 ----TGTTVAASADGSRFWAPGDPQPV--VYAVGFNSWAASQGVPANAIQIRSDRVNP 579  
 Db 1658 IDPSNGSVIAGTAEGATVILTIDGNGNPIQVTVADSGN--WSFTPGTP----- 1704  
 Qy 580 KTFVALSNGTFYR-----STDGGV--TFQPVAAAGLP---SSGAVGVMPHAPVKGED 626  
 Db 1705 ----LSNGTVVNAVAQDAAGNTSGPVSTTVDAVAPATVIDPSNGVELSGTAEPGRVI 1759  
 Qy 627 LMLAASGLYHS--TNGSSSIAITGV-----SSAVNVGFGKSAPGSSYPVAVFWGTIGGVT 681  
 Db 1760 LTDGNGNPIQOTLADSGNWSFTPGTFLANGTVVNAVAQDPAGNTSGPASITVDTVAPAT 1819  
 Qy 682 GAYRSDCGTTWILNDQHOYGNWGAITGDHANLRVYIGTNGRGIVYDIDGAPSG 740  
 Db 1820 -----PVINPSN-----GSVITGT-AEVGAKVILTIDGNGNPIGETTADGSG 1859

## RESULT 11

ABP59933  
 ID ABP59933 standard; protein; 2468 AA.

AC ABP59933;

DT 28-AUG-2003 (first entry)

XX Microbial resistance gene PA1874 protein.

DE Biofilm; microbial resistance.

KW Unidentified.

OS WO2003041483-A2.

PN 22-MAY-2003.

PD 18-SEP-2002; 2002MO-US029565.

XX 18-SEP-2001; 2001US-0323241P.

PR (DART-) DARTMOUTH COLLEGE.

PA O'toole GA, Mah T;

PI WPI; 2003-468567/44.

XX N-PSDB; ACC59398.

XX

PT Identifying modulators of microbial resistance of organisms in biofilms,  
 PT e.g. inhibitor of biofilm formation, by employing expression controls, or  
 PT efflux pumps containing polypeptides, of genes associated with biofilm  
 PT resistance.  
 XX  
 PS Claim 33; Fig 7; 102pp; English.  
 XX

CC The present invention relates to a method of identifying a compound  
 CC capable of altering the sensitivity of a microorganism to an  
 CC antimicrobial agent by employing efflux pumps comprising polypeptides  
 CC encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143,  
 CC PA2390 or PA1163. The method is useful for identifying modulators of  
 CC microbial resistance of an organism in a biofilm. The methods are also  
 CC useful for identifying genes that encode proteins that play a role in  
 CC biofilm resistance. The method is particularly useful for screening  
 CC compounds or discovering compositions that will inhibit biofilm formation  
 CC and overcome their resistance mechanisms. These methods are particularly  
 CC useful in medical, industrial or natural settings, where formation of  
 CC biofilms can have serious negative consequences and result in high costs  
 CC both in human health and economic terms. The present sequence is a  
 CC protein shown in the exemplification of the invention

SQ Sequence 2468 AA;

Query Match 5.9%; Score 238; DB 6; Length 2468;

Best Local Similarity 22.8%; Pred. No. 4.8e-07;

Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

Qy 11 VAIGGGFVDGIVFNEGAPGLIYVTDIGMYRDAANGRWIPLLDWGNWNGY----- 65

Db 1084 VNLSNGSSLSG---TAEFGSTVILT-----GNGNPIAEVTADSGNWTYTPSTP 1130

Qy 66 --NG-VVSIADPINTNKVAAVGMVYNS-----WDPNDGAILR--SSDQATWQITPLP 115

Db 1131 IANGTVVNVVQAQDSGNSPPATVTVDSAPPAPVNPNSGVVISGTAEAGATVLT--- 1187

Qy 116 FKLGNNMPGRGGERLAVDPNNDNLIYFGAPSGKGLMRSTDGATWSQMTNFPDVGTYIA 175

Db 1188 -DAGNPIQVTVAD-----GSGNWSFTP-----GTPLA 1214

Qy 176 N-----PTDITGYQS---DIQGVVVAFAFKSSSL--GQASKTIFVGVADPN-N 218

Db 1215 NGTVIVATATPTGNTGPOAATTVDAVAPPAPVIDPSNGTTISGTAEAGAKVILTIDGNGN 1274

Qy 219 PV-FWSRDGGATWQAVPGAP-----TG-----FIPHKGVDPPVN 251

Db 1275 PIGETADSGNWSFTPGTFLANGTVVNAVAQDPAGNTPQGSTTVDAVAPNTPVNPSPN 1334

Qy 252 HVLVIATNTGGPY---DGSSGDVWKFVSVT-SGTW--TRISPVPS-----TDTANDYFGYS 301

Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIQQTADSGNWSFTPGSQLPNGTVVNVVTSADAAGNT 1394

Qy 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDDG-----ATWTRIDW 342

Db 1395 SLPATTVDSSLSPIQVDPNSGVSIGTADAGNTIII--TDGNGNPIQVTVADSGNWS 1452

Qy 343 WT-----SYNRSRLRYVLDISAEPLW--TFGVQPNPPVPSPKLGWDEMAIDPFNS 392

Db 1453 FTPGIPLPDGTVVNVVARSNSVDSAPAVITVDGAPAPV-----IDPSNG 1499

Qy 393 DRMLYGT---GATLYATNDLTWKDSGGQIHIAPMVKGLEETAVNDLISP---PSGAPLI 445

Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIQATADSGNWTFTPTPLANGTVIN 1548

Qy 446 SALGDLGGFTH-----ADVTAVPSTIF--TSPVFTTGTSDVVAELNPSIIVRAGSPDPS 497

Db 1549 AVAQDPAGNTSGPASVTVDAIAPPAPVNPNSGVVISGT----AEAGATVILTIDGNGNPI 1604

Qy 498 SQPNDRHFAFSTDGKNWF-----QSEPPGGVT----- 525

Db 1605 GQ-----VTADSGNWSFTPGTFLANGSVINALAQDAAGNNSPTSATVDSLAPAPV 1657

Qy 526 ----TGTTVAASADGSRFWAPGDPQPV--VYAVGFNSWAASQGVPANAIQIRSDRVNP 579



Db 1658 IDPNSGVIACTAAGATVILTDGNGNPIDQVTDGSGN-WSFTGTP----- 1704  
 Qy 580 KTFYALNSGTFR-----STDGCV--TFQPVAAAGLP---SSGAVGVMFHAPGKEGD 626  
 Db 1705 -----LSNGTVVNAQAAGNTSGPVSTTVDAVAPATFVIDPSNGVLSGTAEFGVRV 1759  
 Qy 627 LMLAASGLYHS-TNGSSWGAITGV-----SSAVNVGFGKAPGSSYPVAVFVGIGVT 681  
 Db 1760 LTDGNGNPICGTLDAGSGNWSFTPTPLANGTVVNAQAQDPAGNTSGPASTVDTVAPAT 1819  
 Qy 682 GAYESDDCGTTLVINDDQHQYGNWQAITGDHANLRVYITGNCRGIVYGDIGCAPSG 740  
 Db 1820 -----PVINPNS-----GSVITGT-AEVGAKVILTDGNGNPICGETTADGSG 1859

## RESULT 12

ABU19529  
 ID ABU19529 standard; protein; 2435 AA.

XX AC ABU19529;  
 XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #5056.

XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX KW Borrelia cepacia.

XX OS WO200277183-A2.

XX FN 03-OCT-2002.

XX PD 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-ESDB; ACA23399.

XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX PS Claim 25; SEQ ID NO 47453; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC on a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by this  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2435 AA;

Query Match 4.9%; Score 199.5; DB 6; Length 2435;  
 Best Local Similarity 23.0%; Pred. No. 0.00022;  
 Matches 185; Conservative 72; Mismatches 303; Indels 243; Gaps 42;  
 Qy 47 ANGRW-----IPLLDWGVNNWYGVVSIADPI--NTNKVWA----- 83  
 Db 173 ANGRWSYPTGTLPEGAHQITVTQDSSGNTSELSTVGI VDTTIVQKTPATAVSDDAQHP 232  
 Qy 84 -AVGMYTNSWDPN-----DGAAIL---RSSDQGATWQITPLPKLG----- 119  
 Db 233 VTIGGATSDTTTLGCTAEAGSRVSVYDGTLLGTTTADSSGKWTFTPTT-GLGEGAHSI 291  
 Qy 120 -----GNMPCRGMERLAVDPNDNLTLYFGAPSGKGLWRDTSQATWSQMTN----FP 168  
 Db 292 TVTATDAAGNVSTPSAAFELTIDTAPALFTVNATDGTSLSGTAEAGATVNIDTNGDGP 351  
 Qy 169 DVGTVIANPTDTGYQSD---IQGVVWV-APEKSSSLGQASKTIFVGVADPNPNPFW 223  
 Db 352 D-ATVADPSGVTYTPSTPLPIGTIVGTATDAAGNTGPSASVTVTGDTTAPGAPVI-- 408  
 Qy 224 RDGGATWQAVPGAPTGFPHKGVFDPVNHVLYATISNTGCP---YDGS-----SG 270  
 Db 409 --GTVTDDA--GSVVGAIASGGSTDDATPTL-SGTAEAGSTVSVYDGTLLGTTTADPSG 463  
 Qy 271 DVWKFSTVSG--TWTRISVPSTDTANDYFGVS---GLTIDROHPNTIMV-ATQISWPD 324  
 Db 464 N-WTFTPTTGLEGAHSLTVTATDTAGNVSVSTAPDLDITDTAPSIPTVNATD-----G 517  
 Qy 325 TIFRSTDCGATWTRINDWTSYPNRSLRYVLDISAPWLTFFGVQPNPPVPSKLGWDEA 384  
 Db 518 TSLSGTAEAGATINIDTNGDGPDAT-----VTADPGGAWTVPSTPLFA----- 562  
 Qy 385 MAIDPFNSDRMLYGTGATLYATNDLTWKDSGGQIHAPMVKGLEEAVNDLISPPSGAPL 444  
 Db 563 -----GTVIGVTAT-----DAAGNTGPSASV-----TVTGTDTAP--CAPV 596  
 Qy 445 ISALGDLGGTHADVAVPSTIFTGFTTGTSDVYAEINPSIIVRAG-----SFDPS 497  
 Db 597 ICTVTDDAGSVVGAIASGGSTDDATPTL--SGT-----AEAGSTVSVYDGTLLGTTTADPS 651  
 Qy 498 SQPNDRHVAFTDGGKNWF-----QCSEPGCVT---TGGTVAASADGSRFVWAGDP 546  
 Db 652 G-----NMTFTPTTALGEGAHSITVTATDTAGNVSVSTAPDLDITDTAP 696  
 Qy 547 GQPVVYAV-GFCNSWAASGVGPANAIQRSDRVNPKTFYALSNGTFFYRSTDGGVTFQPVAA 605  
 Db 697 ALPTVNATDGTSLSGTAEAGATVNIDTNGDGPDTATVTDPSGAW-----TYTP-ST 747  
 Qy 606 GLPSSGAVGVMFHAPGKEGDLWLAAASSGLYHSTNGSSWSAITGVSSAYNVGFGKAPG 665  
 Db 748 PLPAGTVIGVTATDAAGN-----TGFSASVTVTGDTTAPG 782  
 Qy 666 SSYPVAVFVGT-----IGGVGTGAYRS-----DDCGTTWVLINDDQHQYGNWQAITGDHANL 717  
 Db 783 AP-----VIGTVDVGVSVGAIIISGGSTDDATPTLS-----GTAEAGSTVS- 824



```

Qy 718 RRVVIGTNGRGIVYDGGAPSG 740
Db 825 --VVDGTTLLGTTAD----PSG 841

RESULT 13
ADA34216
ID ADA34216 standard; protein; 3892 AA.
XX
AC ADA34216;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #1377.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA30090.
XX
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
XX Example; SEQ ID NO 5503; 328pp; English.
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX The A. baumannii nucleic acids and polypeptides are useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, to detect the presence of
XX A. baumannii and other Acinetobacter species in a sample, in screening
XX compounds for the ability to interfere with the A. baumannii life cycle
XX or to inhibit A. baumannii infection, and as biocontrol agents for
XX plants. The present sequence represents the amino acid sequence of an A.
XX baumannii protein.
XX
SQ Sequence 3892 AA;

Query Match 4.5%; Score 181.5; DB 6; Length 3892;
Best Local Similarity 21.5%; Pred. No. 0.0072;
Matches 204; Conservative 90; Mismatches 344; Indels 309; Gaps 54;

Qy 7 TWSNVAICGGGFVDG-----IVFNEGAPGILYVR-TDIGMYRWDAAANGRIPLLDW 57
Db 543 TWT-VSVPGSLVADADKTDIAKVTFTDAAGNSSTVNDTQI---YTLDTA-APAAPVIDP 597
Qy 58 VGRNNGYNGV--VSIADPINTNKVWAAVGMYTN-----SWD-PNDGAILRS 102
Db 598 V-----NGTDPITGTAEFGSTVTV-----TYPNGDTATVVGPDGSGWSVPNPG--LND 643
Qy 103 SDQATWQITPLPKLGN--MPCRGMGERLAVDPNN-----DNILYFGAPSG 148
Db 644 GDEVEAIATDP-----AGNPSLPGTATVD--AVGNTDGVNFTVDSVTADNVINASEASG 696
Qy 149 K----GLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSSLGQ 204
Db 697 NVTYGV/LKNVPADAANTVTVVINGQTYTATVDSAG-----TWTVSVPGSLDALT 748

```

```

Qy 205 ASKTI-----FVGVADPNPNVFW 222
Db 749 ADKTIADAKVTFTDAAGNSSVNDTQYTTIDTTAPDAPVINPVNGTDPIGTAEFGSTTV 808
Qy 223 SRDGGATWQAVPGAP-TGFIPIHKGVDPVNHVLYTATNTGPGYDGSDDGVKFSVTSCT 281
Db 809 TYPDGSITTVVAGPDGTWTVNPGNDG-DKVTATATDPAGNP-----SLPGT 855
Qy 282 WTRISPVSTDTANDYFGYGLTIDR-----QHPNTIMVATQI 319
Db 856 ATVDAGVNTDGVN--FTVDSVTADNVINASEASGNVTVTVGLKNVPADAANTV-VTVVI 912
Qy 320 SWPDPITIIFRSTDCGATWTRIMWTSYPNRSRYVLDISAEPMLTF----- 365
Db 913 NGQTVTATVDSITAG--TWT---VSVPGSLDADADKTIDAKVTFTDAAGNSSSVNDTH 965
Qy 366 -----GVQPNPVPSPKLGWDEAMAIIDPFNSD-----RMLY--GTGATLYATND--- 408
Db 966 TTVTDTVAPNAPV-----LDPINATDPVSGQAEFGSTVTVTPDGTATTAVAGPDGSW 1018
Qy 409 -----LTKWDSGGQIHI-----APMVKGLEETAVND-----L 435
Db 1019 SVNPNGLVDSGTVTATATDPAGNTSLPGTGTVSADITAPVV-ALDDVLTNDSTPALTGT 1077
Qy 436 ISPPSGAPLISA-----LGDLGGFTHAD-----VTAVPSTIFTSPVFTTG----- 475
Db 1078 VNDPTATVVNVVDGYDPAVNVGD--GTWTLADNTLPLTDGPHITITVTDAAAGNAGTDT 1136
Qy 476 --TSVDYAEIPLNSIIVRAGSFPDPSQPNDRHVAFTDGGKNWPGQSEPGVTTGTVVAA 533
Db 1137 AVVTIDTTAPNAPVLDPINATDP-----VSGTAEAGST-VTVTYPDGT--ATVAG 1185
Qy 534 ADGSRFVWAPGDPQPV-----VYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGT 589
Db 1186 TDGS---WSVPNPNGLVDSGTVTAT--ATDPAGNTSLPGTGTVSADITAP--VVALDDVL 1238
Qy 590 FVRSDTG--GVTFQVAAAGLPSSGAVGVNMFHAPVPGKEGLWLAASSGLYHSTHSGWSWA 647
Db 1239 TNDSTPALTGTVNDPTATVVN--VDGTDYPAVNVGDG--TWTLDNTLPLVADGPHI-IT 1294
Qy 648 ITGVSAVNVG-----FGKSAPGSSYPVAFVVGITGGVTGAYRSDDC-----GTTW 693
Db 1295 VTATDAAGNAGTDAVVTIDTTAPNA--PVLDPINATDPVSGTAEAGSTVTVTPDGTTA 1352
Qy 694 VLI--NDQHQYGNWQQAITGDHANLRVYIGTN-----GRGIVYGD 734
Db 1353 TVVAGTDGSGWSVPNPNGLVDSGTVTATATDPAGNTSLPGTGTVSADI 1399

RESULT 14
ABU16642
ID ABU16642 standard; protein; 1308 AA.
XX
AC ABU16642;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #2169.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Acinetobacter baumannii.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US0009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.

```

PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA20512.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 44566; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1308 AA;  
 Query Match 4.5%; Score 181; DB 6; Length 1308;  
 Best Local Similarity 24.1%; Pred. No. 0.0018;  
 Matches 162; Conservative 65; Mismatches 245; Indels 200; Gaps 38;  
 QY 150 GLWRSTDSGATWSQMTNPPDVGTVIANPTDTGTQSDIQGVVWVAFDKSSSLGQASXTI 209  
 DB 2 GSWSVNPG-----NLVGDGTATATDPAG-----NTSLPGTGVSV--- 38  
 QY 210 FVGVDNPNVFWSRDGGATQWQVGPATGTI--PHKGVPDPVNHVLYATSNVTGGPYDG 267  
 DB 39 ----ADITAPVV-ALDDLVLNDSTP-ALTGTVNDPTATVWVNDGVDYPAVNN----- 85  
 QY 268 SSGDVWKFVSTGWTWTRI-----SPVSTDTANDYFGYSG-----LTIDRQHNPTI 313  
 DB 86 --GD-----GIWTLADNLTPLADGPHITVTATDAAGNVGNDTAVVTDITVAPNAP 135  
 QY 314 MVATQISWMPDTI-----IFRSTGGATWRIWDTWSYVNSRLRVLDISAEPMLTFGVQ 368  
 DB 136 VL-----DPINATDPVSCQAPGSGTIVT-----VTYPDGTATVAGTGSW----- 176  
 QY 369 PNPVPSPKL---GWMDEAMAIDPNFSRMLYGTG-----ATLYATNDLTKWDS----- 414

DB 177 ---SVNPNGLVDGDTVTATATDP-AGNTSLPOTGTGTVSADITAPVVALDDVLNDSTPAL 232  
 QY 415 GGOIH-----IAPVMKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTS 469  
 DB 233 TGTVNDPTATVVNVVDGVDYPAVNN-----GD-GTWTLADNLT--PTLADG 275  
 QY 470 PVFTTGTSDVYAEI--NPSIIVRAGSFDPSQPNDRHVAFSTDGKKNWFGSGPGGVT-- 525  
 DB 276 PHTITVTATDAAGNVGNNDTAVVTDITVAPNAPVLDPINATDPVSGQ-----AEPGSGTIV 330  
 QY 526 -----TGGTVAASADGSRFVWAPGDPGPV-----VYAVFGNSWAAASQGVPAQAIRSDR 576  
 DB 331 TVPDGTTATVAGTDS-----WSVNPENGLVDGDTVTAT--ATDPAGNTSLPGTGTVSADI 385  
 QY 577 VNPKTFYALSNGTFYRSTDG--GVTFQPVAAAGLPSGCAVGMFHAVPGEGDLWLAASSG 634  
 DB 386 TAP--VVALDDVLNDSTPALTGTVNDPTATVVVN--VDGVDYPAVNNGDG--TWTLADNT 440  
 QY 635 LVHSTNGSSWGAITGVSSAVNVGFKSAPGSSYPVAVVGTGTGGVTVGAVRSDDC----- 689  
 DB 441 LPTLADGPHT-ITVTATDAAGNVGYD-----TAVVTIDTVAPKCTGAXSQCNRLQX 491  
 QY 690 -----GTTWVLINDDHOYGNW-----GQAITGDHANLRVVGITNGR-- 727  
 DB 492 AVRLEPGSTVTVVTPDGTATVAVDTPD-----GSMVSNPNGLVDGDTCDLSQSYXSUGNT 548  
 QY 728 -----GIVVVDI 734  
 DB 549 SLPGTGTVSADI 560  
 RESULT 15  
 ABU22692  
 ID ABU22692 standard; protein; 1074 AA.  
 XX AC ABU22692;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by Prokaryotic essential gene #8219.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Burkholderia mallei.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA26562.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 50616; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression

of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide(s) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

xx SQ Sequence 1074 AA;

Query Match 4.2%; Score 170; DB 6; Length 1074;  
 Best Local Similarity 21.6%; Pred. No. 0.008;  
 Matches 186; Conservative 82; Mismatches 300; Indels 292; Gaps 44;

Qy 10 NVAIG--GGGFVDGIVNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVMNNWYNG 67  
 Db 180 NVGLGQAGQFVTGT--GNSAMGHLAGST-VSGSY--NAAFG-----EYAGTNTSG-- 225

Qy 68 VVSTADPINTKVAAGVMTNSW-DPNDGAILRSSD---QGATWQITPLPFLKGNMP 123  
 Db 226 -----GAN---AAGFGYAGRYINGTNTALGAYDLPVNGTWYGS---YVTGSNNL 270

Qy 124 GR-----CMGERLAVDPNNNLYFGAPSGKGLWSTDSGATWSQ---MTN 166  
 Db 271 GAGHNSGAYVSGASNVGLGDGAGFTVGSNNVAITGTAAGSGAYTSGPGSATLNAALVASN 330

Qy 167 FPDVGTIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDG 226  
 Db 331 TVSIG-----TRATASQSDA-----IAGKGATASGAQSISI----- 362

Qy 227 GATWQAVPGAPTGP IPHKGVDPVNNHLYIATSN TG GPYD--GSSGDVWKFSVTS GTWTRI 285  
 Db 363 -GTGNVSVSGKSG-----AIGDP-----STVSGAGSYSIGNNTV-----ANSNTFVLG 405

Qy 286 SPVPSTDTANDYFCYSLGTLTD RHPNTIMVATQISWHPDTHIIFRSTGGATWTRIDWTS 345  
 Db 406 NGVTTT-----QDNSVVLGNQSTDRAAVAVSSETINGTTY-----N 441

Qy 346 YPNRSLRYLVDISAEPWLTFCGVQNPVPVPPKLGMDMAIDPPFNDRMLYG----- 398  
 Db 442 YAG-----VASPANG-----VVSIGGVGTERQLINVAAGQVS 473

Qy 399 -----TGATLYATNDLTKWDSGGQIHIA PMVKGLEETAVNDL----- 435  
 Db 474 ATSTDAINGSLYATN-----QAVIA-----EDAKVNSLGGGVASALGGNAAYNA 518

Qy 436 -----ISPPSGAPLISALDGLGGFTHADVAVPSTIFTSVPTTGTSDVYAE LNPSSIIVRA 491  
 Db 519 TTGATITAPSYAVVGTTONSVGGAI DALQALPLOYTSGPGVTTTPNAPGSAPTNTVTLVGA 578

Qy 492 GSPDPSSQP-----NDRHVAFSTD-----GCKNHPQSGSEP 521

Db 579 GPGGANTTPVTLTNPAPGKL SATSTDAVNGSOLYATNQVANLVSSVNNGGVGPVQYSDP 638  
 Qy 522 GGVTT--GG-----TVAASADGSRFVWAPGDPGQPVVYAVFGNSWAAASQGVYAPNAQIR 573  
 Db 639 SAPITPNGGKPSQDLTLVGAASGPVALHNVA-PGTASTDAVNVGOLGAVTTGLGGGAAI- 696  
 Qy 574 SDRVNPKT-----FYALSN--GTFYRSTDGGVTTQPV-AAAGL-----PSSGAVG- 614  
 Db 697 ----DPKTGAVTAPSYTYVNADGTTNSVSNVGAADAINSTGIKYFHANSTKPDQALGA 752  
 Qy 615 ----VMPHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVS-----SAVNVGFGK 661  
 Db 753 DSAVIGPNAVANNAGD--VALGSGAVTSOAGTSLSETINGVTYSFAGTTPIGTVSVG--- 807  
 Qy 662 SAPGSSYPAVFVVGITGVGTGAYRSDDCGTTWVLINDDQHQYGNWG-QAITGDHANLRRV 720  
 Db 808 -APGVERT---ITNVAAGRIQSSSTD-----ALNGSOLYGTNQSI EALTDKNSLGNT 856  
 Qy 721 YITNGRGIVYDGI GGAPSG 740  
 Db 857 VANTLIGSGASYNPQTGA VNG 876

Search completed: May 11, 2004, 12:06:49  
 Job time : 125.24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:51:45 ; Search time 34.7001 Seconds  
(without alignments)  
2051.340 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIVYDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97033	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical glyci
6	202.5	5.0	2174	2 B95965	hypothetical prote
7	189	4.7	3972	2 S75231	hypothetical prote
8	186.5	4.6	4199	2 S76412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RTX fami
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 A83528	extracellular seri
17	166.5	4.1	902	2 B87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1684	2 S10789	amylase A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD0835	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

30	157	3.9	809	2 A55547	quinase-shikimate
31	156.5	3.9	2204	2 A70524	probable PPE prote
32	156	3.9	3157	2 B70969	probable PPE prote
33	155	3.8	4180	2 G83559	hypothetical prote
34	154.5	3.8	1032	2 T34433	hypothetical prote
35	154.5	3.8	3570	2 T45025	mucin MUC5B, trach
36	153	3.8	1055	2 A87364	OmpA-related prote
37	153	3.8	1821	2 AG2335	hypothetical prote
38	153	3.8	13288	2 T03099	mucin, submaxillar
39	152.5	3.8	3016	2 S77300	hypothetical prote
40	152	3.8	348	2 T35248	probable oxidoredu
41	152	3.8	1049	2 T42045	beta transducin-li
42	152	3.8	1196	2 A29130	beta-amylase (EC 3
43	152	3.8	3472	2 T31308	hypothetical 367K
44	151.5	3.8	820	2 B72575	hypothetical prote
45	151.5	3.8	1145	2 B75625	hypothetical prote

ALIGNMENTS

RESULT 1

T35237

probable secreted cellulase - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C;Accession: T35237

R;Seesger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z21572

A;Accession: T35237

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-890 <SEE>

A;Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.30C

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC5C7.30C

Query Match 58.7%; Score 2371; DB 2; Length 890;

Best Local Similarity 57.7%; Pred. No. 1.4e-131;

Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY	6	YTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGYRWDAANGRWIPLLDVGWNNWG	65
DB	41	YTKNARIDGGGFVGVFNTEKDLAYARTDIGAYRQEBESHTWTFLLDHVGDWGH	100
QY	66	NGVSIADPINTNKKVAAVGYMNSDPNDGAILRSSDQGATWQITPLPFLKLGNNPGR	125
DB	101	TGVVALASDAVDPDRVYAAVGYTNDWPTNGAVLSADRGASWEKADLPFLKLGNNPGR	160
QY	126	GMGERLAVPNDNLIFYGAPSGKGLWRSTDSGATWSQMTNFPDVGTYTIANFTDTGYOS	185
DB	161	GMGERLAVDPNDNLIFYGAPSGHGLWRSTDSGATWSQMTNFPDVGTYTIANFTDTGYOS	220
QY	186	DIQGVVAVRDKSS--SSLGQASKTIFVGADPNPVFNSRDGGATWQAVPGAPTGFPHK	244
DB	221	DNQGITWTFDESGGGAGTARTIUVGVADKENVYRSTDAWELRAGQGTGLAHK	280
QY	245	GVFPVPHVLIATISNTGPGYDSSGDVWKFSVTSGTWTTRISPFVSTDTANDYFGYSGLT	304
DB	281	GVLDAGNGLYLAISDTCGPGYDGGKGLRYATATGTWTDISPAEADT--YYFGSLT	337
QY	305	IDRQHPNTIMVATQISWWPDYIIFRSTGGATWTRIMWTSYPNRSLRVLDISAPWLT	364
DB	338	VDRQRPQVMTATYSSWWPDYIIFRSTDSGATWSQMTNFPDVGTYTIANFTDTGYOS	397
QY	365	FGVQNPVPPSKLGMWDEMAIDPFNSDRMLYGGATLYANDLTWK--DSGQIHIAPM	423
DB	398	WGAPAPPEQTPKLGWMTAELEIDPFSDRMVYGGATVYGTENTLNWDEGGTFAVEPM	457
QY	424	VKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTCTGSDVDAEL	483

Db 458 VRGLEETAVNDLASPPSGAPLLSALGDYGGFRHTSLTEVPSSMMYTSNFTSTSLDPAET 517

QY 484 NPSIIIVRAGSPDPSSQPNDRHVAFTDGGKWNFGQSEPGVTTGCTVAASADGSRFVWAP 543

Db 518 KPDVVVRAGNL-D-SGP--HIAFTDNGANWFGGTDPSGVSGGTVAAGADGSRFVWSP 572

QY 544 GDPGPVVVAVYVFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGTFFRSTDDGGVTFQPV 603

Db 573 --EGAGVQVTTGFGTSWAQSTGLPAGAVESDRVNPATFYGPKSRFYVSTDDGATFTAS 630

QY 604 AA-GLPSSGAVGVMTHVPGKEGDLWLAASS-----GLVHSTNGSSSAITGVSSAVNV 657

Db 631 AATGLPAGD--GVRFKALPGGEGDVLGGAADGPGYGLWHSTDDGGCTFRLPEVDAADTV 688

QY 658 GFGKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDHOYGNWGAQITGDHANL 717

Db 689 GFGKAAPGASYQLTFTSAEIGGVRGIFRSTADAGATWTRVNDDAHQMGWGTGAITGDPRVY 748

QY 718 RRVYIGTNGRGIVYGDY----GGAPSG 740

Db 749 GRVYVATNGRGIVYGDYSDTGGGTDPG 775

RESULT 2

D37013

probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: D37013

R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: D37013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-839 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:GI5023820; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0919

Query Match 49.8%; Score 2009; DB 2; Length 839;

Best Local Similarity 50.9%; Pred. No. 2.4e-110;

Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;

QY 3 TOPYTSNVAICGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGN 62

Db 37 SQGYKWDNAKIGAGGYVPAVFNKTEKDLIYARTDMGAYRWDKANKWKIPITD--GFSD 94

QY 63 WGYNGVVSIAADPINTNKVAAVGMVMTNSWDNDGAILRSSDOGATWQITPLPKLGGNN 122

Db 95 WTMUGCESIATDPIDITNRVYIAAGLYTNDQDENAYILSSQDKGNTWKRYQPLPKVGGNN 154

QY 123 PGRGMRGLAVDPNNNDILYFAGPSKGLWRSTDSGATWSQMTNPPDVGTYIANPTDITG 182

Db 155 PGRNMRGLQIDPNDKILYLARSNGLWKSSEDYGTWSKVDNFPDYGIVQDQONE-- 212

QY 183 YOSDIQGVVWVAFDKSSSLGQASKTIIFGVAD--PNNPVFWSRDCGATWQAVQVPGAPTGI 241

Db 213 YTAADKGVVWETFDPSGTGKSPSTQTMVGAADTKGNNIYVTDGDKTWSAVKGPQKYL 272

QY 242 PHKGVFDPVNHVLIYATNTGPGYDSSGDVWPKFSVTSQWTRISPVSTDTANDYFGVS 301

Db 273 PHHGL-ASDGLMYSYNTCGPYDSDGQVWYKNTKTGEMTNIPTPAVGDTKS---GFG 328

QY 302 GLTIDRQHPNTIMVATQISWNPDTIIFRSTDCGATWTRIDWNTSPNRSRLRVLDISAEP 361

Db 329 GISVDAQPNVNVVATLNRWPDDEIYRSTDAKTKWKPINDWNGVFNRIYGLNLDYSAQP 388

QY 362 WLTFG-VQPNPVPVFPKLGWMDDEAMAI DPFNSDRMLYGTGATLYATNDLTWKDWSGGQIHI 420

Db 389 WLDWCKTGVTPDDPLVKLGMMGDLLEIDFPNSDRMPYGTGATLYGTDDLTNWDKGNVDI 448

QY 421 APMVKLEETAVNDLISPPSGAPLLSALGDLCGFTHADVTAVPSTIFTSPVFTTGISVDY 480

Db 449 SVKANGIECAVNDVVVPTKGAQLLSAVGDCGCFYHDDITKVPKMTTFFNSATTSIDY 508

QY 481 ABLNPSIIIVRAGSPDPSSQPNDRHVAFTDGGKWNFGQ--GSEPGGVTTGCTVAASADGSRF 539

Db 509 ABSVFNFFVVRVGVNDTSKNQDKCGISYDGGKWNFSAGSNISGVYKAGTVAAGADAKTI 568

QY 540 VMAPGDGPQPVVAVGVFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSGTFFYESTDGGVT 599

Db 569 WSP--EEGANAAYSTDNGKNTWPCSGLPQCAKVRSDRVNPKTFYGFNGFYISTDAGAT 627

QY 600 F-QPVAAGLPSSGAVGVMTHVPGKEGDLWLA--ASSGLYHSTNGSSSAITGVSSAVNV 657

Db 628 FTQSSQTGLPTKKG-GI-FKTVIGHGEDIWAGKDLWHSTDSGATFTKVSVDASDTV 685

QY 658 GFGKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDHOYGNWGAQITGDHANL 717

Db 686 GLGKSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWTRVNDDAHQYGSPDYCITGDPNKY 745

QY 718 RRVYIGTNGRGIVYGDIGA 737

Db 746 GRVFGTNGRGIVYGDIGS 765

RESULT 3

T00349

Avicelase III - Aspergillus aculeatus

C:Species: Aspergillus aculeatus

C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Jul-1999

C:Accession: T00349

R.Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.

submitted to the EMBL Data Library, June 1998

A:Description: Avicelase III from Aspergillus aculeatus.

A:Reference number: Z14141

A:Accession: T00349

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-856 <ARA>

A:Cross-references: EMBL:AB015511; NID:d1199887; PID:d1029971

C:Genetics:

A:Gene: aviii

C:Superfamily: fungal cellulose-binding domain homology

F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 41.6%; Score 1680; DB 2; Length 856;

Best Local Similarity 46.2%; Pred. No. 5e-91;

Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTOPYTSNVAI-CGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59

Db 21 AASQAYTWKQNVVTGGGGFTPGIVFNPSAKVAYARTDIGGAYRLN--SDDTWPLMDWVG 79

QY 60 ---WNNWGVVVSIAADPINTNKVAAVGMVMTNSWDNDGAILRSSDOGATWQITPLPF 116

Db 80 NDTWHDW---GIDALATDPVDTRVYVAVGMVMTNEWDPNVGSILRSTDDGTWTEKLPF 136

QY 117 KLGGNMPGRGMRGLAVDPNNNDILYFAGPSKGLWRSTDSGATWSQMTNPPDVGTYIAN 176

Db 137 KYGGMNPGRGMERGLAVDPNNNDILYFAGPSKGLWRSTDSGATWSQMTNPPDVGTYIAN 196

QY 177 PDITTYQSDIOGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWSRDCGATWQAVPGA 236

Db 197 SSST--YTSDDPGIAWVTFDSTSGSSGSATPRIFGVADAGKSVFKSESDAGATWAWVSGE 254

QY 237 PT-GFIPHKGVDPVNHVLIYATNTGPGYDSSGDVWPKFSVTSQWTRISPVSTDTAN 295

Db 255 POGYGLPHKGVLSPEKTLIYSTANGAGPYDGTNGTVHKNITSGVWTDISP---TSLAS 311

QY 296 DYFGYSGLTIDRQHPNTIMVATQISWNPDTIIFRSTDCGATWTRIDWNTSPNRSRLRVYL 355

Db 312 TYGYGGLSVLDLQVPGTLMVAALNCWPPDELI FRSTDSGATWSPIEWNGVPSINYYSY 371  
Qy 356 DISAEPWLTFCGVOPNP-PVPSPKLGMWDEMAIDPFNSDRMLYCTGATLYATNDLTWKDS 414  
Db 372 DISNAPWIQDTSIQDFPV---RVGMVVEALADPFDSNHWLYCTGLTVYGGHDLTWDS 428  
Qy 415 GGQIHIAPMVKGBETAVNDLISPPSGAPLISALGDLGGFTHADVAVTPSTIFTSPTFT 474  
Db 429 KHNVTVSLAVGIBEMAVLGLITPPGFPALLSAGVDDGDFVHSDLDAPNQAYHTPTYGT 488  
Qy 475 GTSVDVAELNPSIIVRAGSDPPSQPNDRHVAFSTDCGKWFQSGEGVTTGCTGAASA 534  
Db 489 TNGIDYAGNKPSNIVRSGASD--DYPT---LAUSSNFGSTWIADYAAASTGTGAVALSA 543  
Qy 535 DGSRFVWAPGDPGPVVVAVVFGNSWAASQCPVANAQIRSDRVNPKTFYALNSGTFYRST 594  
Db 544 DGDVTLLMSSTSGALVSKSQG---TLTAVSLSPGAVIASDKSNTVYFGSGSAGAIYVK 600  
Qy 595 DGGVTFOFVAAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSSAI-TGVSS 653  
Db 601 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTA 658  
Qy 654 AVNVGFKSAPGSSYPVAVFVGTIGTGVAYRSDCGTTWVLINDDOHQYGNWQA--IT 711  
Db 659 GWSFGFKASSTGSYVVIYGFFTIDGAAGLFPKSEDACTNMQVISEDASHGFGS-GSANVN 717  
Qy 712 GDHANLRVYGTNGRGIVYGDIGGAPSG 740  
Db 718 GDLQYGRVFRGHERPCHLLRQSQREPAG 746

RESULT 4  
F72393  
hypotheical protein TM0305 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: F72393  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: F72393  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-707 <ARN>  
A:Cross-references: GB:AE001712; GB:AE000512; NID:94980799; PIDN:AAD35393.1; PID:9498080  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0305

Query Match 28.8%; Score 1160.5; DB 2; Length 707;  
Best Local Similarity 35.2%; Pred. No. 1.2e-60;  
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

Qy 6 YTSNVAIGGGFVDGIVFNEGAPGILYVVRTDGGMYRWDAAAGRWIPLLDWGNNWGY 65  
Db 21 FEWKSVEINGGGFVPGIIFHPASPGLLYARTDVGLYRWDEETKRWKQLDFLRDQSDY 80  
Qy 66 NGVYSIAADPINTNKVAAVGMVNTNSDPNDGAILRSDQATWQITPLP---FKLGNNM 122  
Db 81 MGVLUSVALDFSDPKRIYAMTKYTDW-AGYGAILISEDYGETWIVNLKYGKIVGNGNE 139  
Qy 123 PGRGMGERLAVDPNNDNLIFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTG 182  
Db 140 DGRNAGERLQVDPNFSVLFMGT-TKYLGLWKSDEFGKWKVDSFPST----- 186  
Qy 183 YQSDIQGVVWVAFPKSSSSLGQASKTTFVGVADPNNPVFWSRDSGATWQAVPGAPTGP 242  
Db 187 -----SVTFVLDFEKGSGKSPTRIFVGCSEPKG-IFVTEDEGTTWNLPLNPLNDLIP 239  
Qy 243 HKGVFDPNVHLYIATSNVCGPYDSSGSDVMKFSVTSQWTRISPVPSPTDANDYFGVSG 302

Db 240 LRGKIH--DGLYVTLNALGNNGATRGAVMKYIADQKQWYDVTMKGD-----FGYCG 291  
Qy 303 LTIDRQHEHTIMVATQISWMPDITIFRSTGDCGATWTRIMDWTSYPNRSRLRYVLIDIGAEPW 362  
Db 292 IDVQE--NVIVSTLDRWPHDEIFISLNGGETWRPILLEKANF-----DINKAPW 339  
Qy 363 LTFGVQPNPPVPSPKLGWMDAMAIDPNRSDRLYGTGATLYATNDLTWKDSGGQIHIA 422  
Db 340 IK--DLNP-----HWISD-VKIDPFDNRRAIFTTGYGVWVTVYELKKSFEF---MGK 384  
Qy 423 MVK-----GLEETAVNDLISPPSGAPLISALGDLGGFTHADVAVTPSTIFTSPTFTGT 476  
Db 385 PVKWIENRGLSEETVVLQVPIGRPLLSATADMGFRHESLDTPPSSMY-KPLKWTSL 443  
Qy 477 SVDVAELNPSIIVRAGSDPPSQPNDRHVAFSTDCGKWFQ-GSEPGGVTTGG--TVAAS 533  
Db 444 GIAFAYQNSKVFARVHTY---TYP---FLSYSEDDGINWREIETVPEGITDGGRLSLAVS 497  
Qy 534 ADGSRFVWAPGDPGPVVVAVVFGNSWAASQGVPA---NAQIRSDRVNPKTFYAL--SNG 588  
Db 498 NDKTLVWSPAN--HEVIVSSDKGSKWKKAISVPVPEFNYFPASDPVNPFSKFYIFDKWNG 555  
Qy 589 TFRSTGCGVTFOFVAAAGLPSSGAVGMFHA-----VPGKEGDLWLAAS-SGLYHSTNGGS 643  
Db 556 DFLISKDGCKGFMK-GAKLPSFDNWNWVSLYSPVLAPDREGDITWLALQWNGLYRSKDGGI 614  
Qy 644 SWSAITGVSSAVNVVGFSGKSPGSSYPVAVFVGTIGTGVAYRSDCGTTWVLINDDOHQY 703  
Db 615 TFERLGNVDIAVYIGFGAPKPGTDYPAIYLNWNGVYVGFIMSTDGKTMWRINNDKHQF 674  
Qy 704 GNWGOAITGDHANLRVYGTNGRGIVYGD 734  
Db 675 G-WHYMIGDMNEFCRIFLTGTEGRGIIVEV 704

RESULT 5  
A83412  
hypotheical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83412  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2468 <STO>  
A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1874

Query Match 5.9%; Score 238; DB 2; Length 2468;  
Best Local Similarity 22.8%; Pred. No. 9.1e-06;  
Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

Qy 11 VAIGGGGFVDGIVFNEGAPGILYVVRTDGGMYRWDAAAGRWIPLLDWGNNWGY----- 65  
Db 1084 VNLSSGSSLSG---TABPFGSTVIITD-----GNGNPFAEVADGSGNWTYTPSTP 1130  
Qy 66 --NG-VVSTAADPINTNKVAAVGMVNTNS-----WDPNDGAILR--SSDQATWQITPLP 115  
Db 1131 IANGTVNVVQAQDASGNSPPATVTVDSAPPAPVNPNSGVVISCTAEAGATVTLT--- 1187  
Qy 116 FKLGNMFPGRGMRGLAVDPNNDNLIFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIA 175  
Db 1188 -DAGGNPIQGVTDAD-----GSGNWSFTP-----GTPLA 1214

QY 176 N-----PTDTTQYQS---DIQGVVWVAFPKSSSL--GQASKTIFVGVADPN-N 218  
Db 1215 NGTVIVATADPTGCTGQAAATVDVAVAPPVADIPDPSNGTTISGTAEAGAKVILTDGNGN 1274  
QY 219 PV-FWSRDGGATQAVPCAP-----TG-----FIPHKGVDFDPN 251  
Db 1275 PIGETTADGSGNWFPTGTPLANGTVNNAVAQDPAGNTGPOGSTTVDAVAPTPVNFPSN 1334  
QY 252 HVLVIATSNTOGPY---DGSSGDVWKFST-SGTW--TRISPVFS-----TDTANDYFGYS 301  
Db 1335 GNLLNGTAEPGSTVTLTDGNGNPIGQTTADGSGNWFPTGSQLPNGTVNVVNTASDAAGNT 1394  
QY 302 GL-----TIDROHPTIMV-----ATQISWPD---TIIIFRSTDDG-----ATWRIWD 342  
Db 1395 SLPATTTVSSLPSPQVDPDPSNGVISGTADAGNTIIII--TDGNGNPICQVTTADGSGNWS 1452  
QY 343 WT-----SYNRSRLRYLVIDISAEPLW--TFGVQPNPPVPSPKLGWDEAMAIIDPNS 392  
Db 1453 FTGPIPLDGTIVNVNAKSPNSVDSPAVITVDGVAPAPV-----IDPSNG 1499  
QY 393 DRMLYGT---GATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISP-----PSGAPLI 445  
Db 1500 TE-ISGTAEAGATVILT-----DGGN-----PIQATADGSGNWFPTPSTPLANGTVIN 1548  
QY 446 SALDGLGFTH-----ADTVAPSTIF--TSPVETTGTSVDYAEIPLNPSIIVRAGSPDPS 497  
Db 1549 AVAQDPAGNTSGPASVTVDALAPPAPVINPNSGVVISGT-----AEAGATVILTDCGNPI 1604  
QY 498 SOPNDRHVAFTDCGKNWF-----QSGEPGGVT----- 525  
Db 1605 GQ-----VTADGSGNWSFTGTPLANGSVINALAQAAGNNSPTSATVDSLAAPAPV 1657  
QY 526 ---TGTVAAASADGSRFVWAFPGDPQPV---VYAVFGNCSWAASQGVPAQAIRSDRVNP 579  
Db 1658 IDPSNGSVIAGTAAGATVILTDGNGNPICQVTTADGSGN-WSFTPGTP----- 1704  
QY 580 KTFVALSNGTYR-----STDGGV--TFQPVAAGLP---SSGAVGVMFHAPVKGED 626  
Db 1705 ---LSNGTVNNAVAQAAGNTSGPVSTTVDAVAPATFVIDPNSGVELSGTAEPGVRI 1759  
QY 627 LWLAASSGLYHS-TNGSSSWAITGV-----SSAVNVGKSKAPGSSYPFAVFGTIGGVT 681  
Db 1760 LTDGNGNPICGTADGSGNWSFTGTPLANGTVNNAVAQDPAGNTSGPASTTVDTPVAPAT 1819  
QY 682 GAYRSDDCGTTWLIINDHOHQYGNWGOAITGDHANLRWRVITGNGRIVGYDIGCAPSG 740  
Db 1820 -----PVINPSN-----GSVITGT-REVGAKVILTDGNGNPICETTTADSGS 1859  
RESULT 6  
E95965  
hypotheical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: E95965  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2174 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAQ49389.1; PID:g15140875; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymb  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation  
C:Genetics:  
A:Gene: Smb21548  
A:Genome: plasmid  
Query Match 5.0%; Score 202.5; DB 2; Length 2174;  
Best Local Similarity 22.7%; Pred. No. 0.00093;  
Matches 176; Conservative 90; Mismatches 312; Indels 197; Gaps 36;  
QY 12 AIGGGGFVDGIVFNEGAPGILVVRT-DIGMYRWDAAANGRWIPLLDVGNWNNWYNGVWS 70  
Db 976 SIGGGG-----GNGGVGSVNSKEISGFNLTA-----NVGVGSGS 1011  
QY 71 IAADPINTKWAAVGMYTNSWDPNDGAILRSDOGA-----TWQITPLPFK 117  
Db 1012 GCASGCGN---AIVGL-----DSGTHLOTSGSGARGVIVQSIGGGGTSCQASVGLS 1059  
QY 118 LGGNWPGRMGRLAVDPNDNLIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANP 177  
Db 1060 ASASLPFGG-EEAAEAEESEEGSAGFSASVSVGRTGGSGS-SGTVNVTTAGT----- 1112  
QY 178 TDTTGVQSDIQGVWVAEPDKS---SSILGOASKTIFGVADPNPVPFWSRDG----- 226  
Db 1113 --ISTFGADADGVLAQSIGGGGLGSGVGOASSGSDSEPLDDEGSECASGNGDDGHGY 1170  
QY 227 -----GATWQAVPGAPTGFIPHKGVDFPNVHLYIATS-----NTGGPYDGSSG 270  
Db 1171 CFGVSVGAT---IDGGTGTGAANGNAVTLTHAGHIATAGDWADGIVAQSIGG--GGAG 1225  
QY 271 DVWKFSVTSGTWTRISVPSTDTANDYFGVGLTIDRQHPNTIMVATQISWPDITIPRS 330  
Db 1226 GTSTAGSGQATANIITGVGGSGAGNGGAGVITFDDNHGNSISTAGSAY--GVLLOS 1282  
QY 331 TDGGATWRIWDVTSYPNRSRLRVLD-----ISAEPWLTFGVQPNPFPVS 375  
Db 1283 IGGGGG---QGGDGSDEAAGRITVGGFGGSGGSGGMVTAKGWINLSTSG----- 1332  
QY 376 PKLGMWDEAMAI--DPFNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHIAPMVKGL 427  
Db 1333 -----DDAHGIVAQSIGGGGVGGAGSSTAEEKHSHTIDLVVGGSG-----VGGSGGE 1382  
QY 428 EETAVNDLISPPSGAPLISAL-----GDLGGFTHADTVAPSTIPTFTSVFTTGTSDY 480  
Db 1383 VDLVSGTTLTS-TSGARALGLVAQSIGGGGIGGAGEAD--SIASLVVGG---SGGCTIDG 1436  
QY 481 AELNPSIIVRAGSFDPSSQPNDRH--VAFSTDGGKN-----WFGSECGGVTTGG 528  
Db 1437 GAVTVDLTSSQSI---TTQGIAAHGLVAQSIGGGGVGGGAAGAPLSFTGNSPSSGDDGG 1493  
QY 529 TVAASADGSRFV---WAPGDPGPVYVAVFGNWSWAASQGVPAQAIRSDRVNPKTFYAL 585  
Db 1494 DVAVTADGSIIFTRGDYAFGLVAQSIGGGGFGGN-ATSAFIGNGNLSSDG----- 1543  
QY 586 SNGTFYRSTGGVTFQPVVA-----AGLPSGAGVGMFHF-AVPGKEGD---LWL 629  
Db 1544 KSGNVTVSLDAGRTQASCKDSIGIFPAQSDAGTDNNGTIDTVNGTGTGSGDNGAGIIV 1603  
QY 630 AASSGLYHSTNGCSWSAITGVSSAVNVGFKSAPGSSYPFAVFGTI-CGVTTGA 683  
Db 1604 SAGKNVTVNNSGNGVSAASGV--AVQYTAGMNSPEDSTLLVNVNAGTITSGSVKGA 1656  
RESULT 7  
S75251  
hypotheical protein slr1028 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S75251  
R:Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis



[illegible]

**RESULT 8**

S76412  
hypothetical protein slr0408 - Synecocystis sp. (strain PCC 6803)  
C:Species: Synecocystis sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
R:Accession: S76412  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
S.

A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4199 <KAN>  
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAA18541.1; PID:g165362c  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: Synecocystis hypothetical protein slr0408

Query Match 4.6%; Score 186.5; DB 2; Length 4199;  
Best Local Similarity 19.2%; Pred.No. 0.019;  
Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

Qy	9	SNVAGGCGFDGV-----FNEGAPGLYVRTDIGMYRWDAANGRIPLL 55
Db	2261	ADVASNGFVIDGNLIGNPPTTFETTSQYDITTPAILINGSNLAYKGFQGNQ---IY 2317
Qy	56	DWGMNN-WGVNGVVSIAADPINTNKVM-AAVGYMTNSW----DPNDGA-ILRSSDQA 107
Db	2318	FTVTSTNGQSNSEVOL---FQSAQTIPPAIAFPENNVLYLAYVDGNGLNIITSODQG 2374
Qy	108	TWQITPLPFKLGGNMPG-----RG-MGERLAVDPNDNIIYFGAPSKGLWRSTDGA 159
Db	2375	TWN---APALAGTSTPTTLFVQGTLSLLFAANNSTSVLFQYLNSNEWIYANEIGS 2431
Qy	160	TWSQMT-----NPPDVGTYIANPTDTTGYSODIQGVVVAF----- 195
Db	2432	NQTAISALSATVLGDTLYLVKGGRNTVPSTLDYITSTN-----ADLSANDWSSPIPG 2486
Qy	196	--DKSSSSLGQASKTI FVGADPNPV-FWSRDGGATQWAVEGAPTGI---PHKGVFDP 249
Db	2487	VSSQGFSLTNDGTNLVLSYLDSSNQNFVSGENGINWSS-PQVITNNISQSPPAIAF-- 2543
Qy	250	VHVLYIATSNTGGPYDGSSGDVWKFSVTGWTRISPVPST-----DTAN 295
Db	2544	ANNELYLS-----YPGQGS-QELNVTS-----FPLPFTGISLGNSLVRFLGDVNG 2589
Qy	296	DYFG--YSGLTDRQHPNTIM-----VATOISWWPDITIFIRSTDGGATWIRWDWTSY 346
Db	2590	DGFADVFGGT-----NAGAIFGNSTKDLLTTASGEDLVI-----SV 2628
Qy	347	PNRSIRVYL---DISAEPLWTFGV-----QPNNPPVPSPKL 378
Db	2629	PNAITRDVISVGDENGDDGKDGLVDGNGFVVVLGNTSLGDLKTLISITSSSPVIVNQ 2688
Qy	379	GWMDEAMAI DFNDSR-----MLYGT-----GATLYATNDLTKWD----- 413
Db	2689	GGVTKSMAGIDYNGDGYDDVLLHWDGNGNQVAWGNSTGVLSNFTNIDYPETOTATTGVDLN 2748
Qy	414	-----SCGQIHAPVKMGLEETAENVDLISPPSGAPLISALGDLGGTHAD 458
Db	2749	SDGIPEITAIGDERKIACQISTSGSFSLPTPTTSSSVINTLAANAOLENIGDFNGDCIAD 2808
Qy	459	VTAVPSITFTS-----PVFTTGTSVDYAELN-----PSIIVRAGSFDFS 497
Db	2809	LAVLASNYAAAGIEPNLPVNLSPGNOGVFI FYGNSGNLSNTAQPDVILAAPPNTFS 2868
Qy	498	SQ-----PNDRHVAFSTDTGGKNW----- 515



Db	2869	QGI <sup>ST</sup> YQL <sup>S</sup> RIAQAGDVNGDGFDDLIS <sup>PT</sup> YVDAENNQG <sup>GV</sup> VFVFGDGDWNNQ <sup>PF</sup> DLGQ	2328
Qy	516	-----FQSGP <sup>GG</sup> VTG <sup>TT</sup> GTVAASA-----DGS <sup>RF</sup> -----VWAP <sup>G</sup> -----	544
Db	2929	LRANQ <sup>SG</sup> SNR <sup>PR</sup> FAIDGSPNSQAGIALNGGGDINGGDFADFI <sup>IG</sup> APENN <sup>LV</sup> QNO <sup>IV</sup> F	2988
Qy	545	-----DPGQ <sup>PV</sup> YAVG-----FGNSWAASQ <sup>GV</sup> PANAQI <sup>RS</sup> DRV <sup>N</sup> -----PKTFYAL	585
Db	2989	IENGELSD <sup>DD</sup> KYGIYIL <sup>LD</sup> GNQTIQMGGDWQANQ-VW <sup>TN</sup> -QVATN <sup>WN</sup> SSR <sup>RP</sup> EAVIGQ	3046
Qy	586	SNGT <sup>FY</sup> SRST <sup>DG</sup> GVTFQ-----PVAAGL <sup>PSS</sup> SGAV <sup>GM</sup> FHADV <sup>PG</sup> K	623
Db	3047	SGNDI <sup>WY</sup> PGCNQ <sup>W</sup> SGWGL <sup>PA</sup> INELAV <sup>NW</sup> NTSG <sup>NP</sup> QIIAGL <sup>GK</sup> G--GIEVY-----	3099
Qy	624	EGDLWLAASGL <sup>VH</sup> STNGSSW-SAIT <sup>GV</sup> YSSAV <sup>NV</sup> FGK <sup>SA</sup> PSS <sup>VP</sup> AVF <sup>VG</sup> TIG <sup>GV</sup> VTG	682
Db	3100	NGT <sup>W</sup> -NNGP <sup>TQ</sup> -----GGW <sup>RS</sup> AT <sup>OM</sup> A-----VOMG <sup>ED</sup> --GS--PSQI <sup>VV</sup> VL <sup>GD</sup> AG <sup>AV</sup> I	3145
Qy	683	AYRSD <sup>CG</sup> TT <sup>WL</sup> IN <sup>DQ</sup> HO <sup>Y</sup> GNW <sup>QA</sup> ITG-----DHANL <sup>RR</sup> V <sup>IG</sup> TN	725
Db	3146	Y <sup>NTQ</sup> -----S <sup>W</sup> RTIN-----NFGK <sup>SV</sup> TL <sup>QS</sup> VOMQ <sup>EA</sup> SNP <sup>NI</sup> VGL <sup>DN</sup> SEV-QY <sup>YQ</sup> GSN	3194
Qy	726	GRGIVY <sup>GD</sup> IG 735	
Db	3195	GV <sup>WTQ</sup> HD <sup>DG</sup> 3204	
RESULT 9			
T13256			
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T			
C/Species: Lactococcus lactis phage BK5-T			
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999			
C/Accession: T13256			
R/Boyce, J.D.; Davidson, B.E.; Hallier, A.J.			
Appl. Environ. Microbiol. 61, 4089-4098, 1995			
A/Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5			
A/Reference number: Z17646; MUID:96064422; PMID:8526523			
A/Accession: T13256			
A/Status: Preliminary; translated from GB/EMBL/DDBJ			
A/Molecule type: DNA			
A/Residues: 1-1904 <BOY>			
A/Cross-references: EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAAA98579.1			

	Query Match	4.5%; Score 182; DB 2; Length 1904;
	Best Local Similarity	20.7%; Pred. No. 0.012;
	Matches	Conservative 83; Mismatches 260; Indels 380; Gaps 55;
QY	1	ATTQPWTWSNVAIGGGFVDGI VFNEGAPG-----ILYVRTDIGMYRWDRAANGR 50
DB	83B	AKPSDTWYS--LIRGNDGKDGATGKGVGAGKGVGIKTVTITYALSSSG---TDRPNTG 891
QY	51	W---IPLLDVWGNNNGYNGVWISIAADPINTNKVAAAVGMYTNSWPNDGAILRSSDOGA 107
DB	892	WTSQVPFL-----VKQYLWKTKVWT---YTDUS-----SSETG- 921
QY	108	TWOITPLPKLGMMPCRGMRGLAYDPNNNDNILYFGASPGKLWRSTDSGATW-SQMTN 166
DB	922	-YSVTVI--AKDGNNGNDGIAGKGVGIKTTITYAVGTSG----TTAPASCWNVSQVPN 973
QY	167	FP-----DVCTYIANPTDDTYGOSDTCQVVWAFAFDKSSSLGOASKTI FVGVA DPNN 218
DB	974	VPAQGFWLTXTVWTVTDN-TSETGYSAMVMGVKDGDGPNGNTN-----GIA--- 1020
QY	219	PVFMSRDG-----GATWOAVPG--APTG-----FIP--HKGVF-----DP 249
DB	1021	----GXDKGKI KATALITYQASPNGTAGTGATWSAPPVAKGSFLMTRIIWTVDMTNT 1076
QY	250	VNHVLVIATSNTGCGPYDSSGDWWKFPSVTSQTWTRISPSPFTDTANDYFCYSGLTIDROH 309
DB	1077	GYAVALMGINGNNG-HDGFFG-----KDGTGIKTTIT-----YAGTSGTTP 1118
QY	310	PNTIMVATQLSWHPDITIIFRSTDGGATWTR-IWDWTSYPNRSLRYVLDISAEPMLTFGVQ 368

Db	1119	PN	-----NGWTSTVP-TVAEGNYLWTKVTWITYTD--NTS-----	1149
Qy	369	PNPPVPSPKLGWDEAMAI	-----DPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHAPM	423
Db	1150	-----ETGYSVAMGVKGDGPNN-----	-----GTNGTAGDKG-----	1180
Qy	424	VKLEETAVNDLISP-PSGAPLISALGDLGGFTHADVAVP--STIFTSPVFT--	-----	473
Db	1181	-KGIKATAITYQASPNGTTAPT-----	-----GTWSASVPPVAKGSFLWTRTIWTYTDNTTE	1231
Qy	474	TGTSVDYAEIN-----	-----PSIIIVRAGSFDPSSQPNDRHVAFTDGGKNWF	516
Db	1232	TGVAVATMGNGNNGHDGPFPGKDGTKIKTTIITYAGSTSGTTPNN-----	-----GW-----	1279
Qy	517	QGSEPGGVTTGGTVAASADG-----SRFVW-----	-----APDGPQPVVVAVG	555
Db	1280	-----TSTVPTVAEGNYLWTKVTWITYTDNTSETGYSVAMGVKGDGP-----	-----	1323
Qy	556	FGNSWAASQGVPANAIQRSDRVNPKTFYALNSGTFFRSTDDGGVTFQPVAAAG	-----	606
Db	1324	-GNN--GTNGI-AGKDGKGIKATAITYQASPNGTTAPTCTGWSASVPPVAKGSFLWTRTIW	-----	1379
Qy	607	-----LPSSGAGGVW-----	-----FHAVPGKEG-----	644
Db	1380	TYTDNTETGYAVAYMGTCNGNCHDGPCKDGTGKIKTTIITYAGSTSGTTPNNGWTSTV	-----	1439
Qy	645	-----W-----	-----SAITGVSSA-VNVVGKSAFGSSYPVAVFVGTIGGVTG--	682
Db	1440	PTVAEGNYLWTKVTWITYTDNTSETGYSVAMGVKGDGPNN-----	-----GT-NGIAGKD	1491
Qy	683	-----AYRSDCGTTWLINDHQGVNKGQAI-----	-----TGDHANLRV-----	720
Db	1492	GKGIKATAITYQASPNGTT-----	-----APTGTWSASVPPVAKGSFLWTRTIWTYTDNTTE	1543
Qy	721	-----YIGTNG	726	
Db	1544	TGVAVAYMGTCNG	1555	
RESULT 10				
F70846				
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)				
C:Species: Mycobacterium tuberculosis				
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999				
C:Accession: F70846				
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; (C				
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Ho				
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.				
Nature 393, 537-544, 1998				
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.				
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complet				
A:Reference number: A70500; MUID:98295987; PMID:9634230				
A:Accession: F70846				
A:Status: preliminary; nucleic acid sequence not shown; translation not shown				
A:Molecule type: DNA				
A:Residues: 1-2523 <COL>				
A:Cross-references: GB:AL021841; GB:AL123456; NID:93261517; PIDN:CAAL17115.1; PII				
A:Experimental source: strain H37RV				
C:Genetics:				
A:Gene: PPE				

Query Match	4.3%	Score 175.5;	DB 2;	Length 2523;
Best Local Similarity	21.8%;	Pred. No. 0.044;		
Matches 175;	Conservative 92;	Mismatches 327;	Indels 209;	Gaps 38;

  

Qy	10	NVAIGGGFVDGIVFNEGA--	PGILYVRTDIDGMRYWDAAN--	GRWIPLLDDWGVNNWGYN	66
Db	1522	NTGIGNSGIASTGLFNAGGPN	TGVVNAVGSYNTGSPFNAGQ	ANTGGFNPGSVNTGWLNTG--	1579
Qy	67	GVVSIADPINTNKKVAAVGM	YNTNSWDPNDDGAILRSS--	DQAGATWQITPLPFKLGGMNPG-	124
Db	1580	-----DINTG-----	VANSQDVNTGAFISGNYNGAFWR	-----GDIYQGL	1614



A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2254

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-908 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAB75287.1; PID:g17132721; GSPDB:GN00179

A:Experimental source: strain FCC 7120

C:Genetics:

A:Gene: alr3588

```
Query Match          4.2%; Score 171; DB 2; Length 908;
Best Local Similarity 20.7%; Pred. No. 0.021;
Matches 161; Conservative 98; Mismatches 272; Indels 246; Gaps 41;

QY 71 IAADPINTNKVAAVGMVNTSWDPNDGAILRSSDQG---ATWQITPLPFKLGNNMPCRGK 127
DB 131 LSAEVIQNNI-----NDNTIANARNIGTENTTITPVVDYVSGS-----171

QY 128 GERLAVDPNNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDI 187
DB 172 GQ--VIDQND---YYS-----FTLTNSGTV-----DIINLSLNGTDT--LYADL 208

QY 188 QGVVWVAFDKSSSLGQASKTIFVGVDPPNPFVWSRDPGGATWQAVP---CAPTGFIPIHK 244
DB 209 QLI-----SSNSVIQTSATV-----GTSLESISRLAAGTYIIRAY 245

QY 245 GVDFPVNHLXI-----ATSNLTGGPYD-----GSSGDVWKFSVT 278
DB 246 SQSDPGNVNLEFNFSADPPDAGGNTSDTSPINLPATFSEIISDOVSLGSDSYQFTLA 305

QY 279 SGTWTRISPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWPDITIIIRSTG-----333
DB 306 SASLVEIQFTSLTADANLY-----LQTONGGNILSTPGCTALDAVRLSLNAGTYNIL 358

QY 334 --CATWTRIWDTWISY-----PNRSL-----RYVLDLSAEPLWIFGV 367
DB 359 VNRGSTETAQYTLGFAQIAGNDQAPNSTTIALNLNLSPISLNEFVGNIDTNDYXPTV 418

QY 368 QPNPVPSPKLGMD-----EAMADP---FNSDRMLYGTGATLYATNDLTKWDS-----414
DB 419 NGTTEIN-----LDLSILNSYLLDPQLVNADVQILNSGGTQVALSNQTSNSESINTIL 472

QY 415 -GGQIHAPWKVGLLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFT 473
DB 473 GAGTYFIRVYTSGLANTFYDLNITAQSOALLVQDINPTGNSDPANLTLGLNTLY---FT 528

QY 474 TGTSDVDAEL---NPSIIIV---AGSPDPSSQ---PNDRHVAFSTD--GCKNW-FQG---518
DB 529 ANDGINGVQLWSSNGDITRILSNISSEFNPTNLIVFNNRLYFAASNDTFGRELWBYNGTV 588

QY 519 -----SRPGGVTTTGGT---VAASADGSRFWA-PGDPGQPVVYAVGFGNSWA 561
DB 589 NRISDINVAGNSNPGNLTVAQNKLFVAVNDVDSIRKLWVNGVNTVNLVDVNASFSNSST 648

QY 562 ASQGVPPANAQIRSRVNPCKTFYALSNGTYFYRSTDG--GVTFQPVAAAGLPSSGAVGVMPHA 619
DB 649 PTFETTTNNQL-----FFTAQNSQLWSTDGTIGTQVISAGGITKSTPRNL---T 696

QY 620 VPGKEGDLWLAAASGLHYSTNGGSSWSAITGVSSA-----VNVGFKSPAGSSYPAVFVG 675
DB 697 VVG--NTLYFTANNG---TSGHEIWQYQNGTASLLEDDITFGNNSFAP--SSLTAV-----745

QY 676 TIGGVTGAYRSDDCGTTWVLNDDOHQYG--NW---GOAITGDHANLRRVYIGTNGR 727
DB 746 -----GNTLYFVTDSDNDNFLELWKSDDGTAAAGTD-----IIGTDGQ 781
```

#### RESULT 13

B85547

probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: B85547

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85547  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5188 <STO>  
A:Cross-references: GB:AB005174; NID:g12513369; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z061  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z0615

```
Query Match          4.2%; Score 170.5; DB 2; Length 5188;
Best Local Similarity 20.3%; Pred. No. 0.22;
Matches 191; Conservative 82; Mismatches 269; Indels 401; Gaps 47;

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDWGVNNWNGVGVVSIADAPINTNKVAAV 86
DB 4083 GANEPAQISTDNGA-----TWVN-----VTVAADSLN-----4109

QY 87 MYTNSWDNDGAILRSSDQATWQITPLPFKLGNNMPCRGMERLAVDPNN-----137
DB 4110 -----WSYVDGRTLTNGT--TTWQVRVV--DIAGNV-GATSSQSALIDTVNPAQVLTIAS 4159

QY 138 -----DNILYF-----GAPSGKGLWRSTDGATWSQMTNFPDVGTYIAN 176
DB 4160 1STDGTSSATDFITSDTWLTLTGLSGLAGLASEVAQISLDSGATWTLTTNGTQWYTD 4219

QY 177 PTDITG---YQSDIQGVVWVAFDKSSSLGQASKTIFGVADPN---NPVFWSRDGGATW 230
DB 4220 RLTLDGSVYQVRV-----LDLAGNTGPVVKTKVVDVTINPTATPTIVSYTDDVG---4269

QY 231 QAVPCAPTGFPHKGVFDPVNHVLIATSNCTGPDG-----SSGDV-----272
DB 4270 -----ORQDTLSLQ-----ATDDTTPLLNGVL SAPLASEGVVLYRNGLLGA 4313

QY 273 -----WKFS---VTSGTWT---RISFPVSTDTANDYFGYGLTIDRQHPNTIMVAT- 317
DB 4314 VTMVGCALNWTYSDSLVSGAYTYSARVVDLAGNITSSDF---VLTVDTSIPTTLAQITS 4370

QY 318 -----QTSW---PDT---325
DB 4371 QTRDTTPIISGVITAALASQYVEWINGKTYTSEPGGVAVVVDPAHNTWYQLPDTDAL 4430

QY 326 -----IIFRSTDGATWTRWDWTSYPNRSRYVL 355
DB 4431 TVSATAYTVTAQVKSSAGNANNANISNGTVTVNAAIDYTPPTWTASKTTAW---GLTYGL 4487

QY 356 DISAEPWLTFGVQPNPPVPSPKLGWMDAMADPFPNSDRMLYGTGATLYATNDLTKWDSG 415
DB 4488 D-SHGWTVLANQVMQSTDP-LTWSKTALT-----LQSGNN-YATSIADYDRN 4535

QY 416 GQIHIAPMVKGLEETAVNDLISPPSGAPLIS---ALGDLGGFT-HADVTAVPSTIFTSP 470
DB 4536 G-----TGDLFITRDDYGTGVINGFTNNGDGT-----FSSA 4566

QY 471 V-FTTGTSVDYAEALNPSIIVRAGSFPDSSQPNDRHVAFST--DGGKNWFGQSEPG-----522
DB 4567 IQVTVGTLTWYGSi-----VAFDKEGDDGYLDFEWIGDAGGPDST 4605

QY 523 -----GYTTGCTVAASADGSRFWAPGDPGQPVVYAVGFGNSWAASQCVPAQAQIRSR 576
DB 4606 FLNNAAGTLVNGSTTSNGGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4655

QY 577 V-----NPKTFYALS-----NGTFFRSTDGCVTFQFPAAGLPSSGAVGVMPHAVPGK-BG 625
DB 4656 VQHTVNLNNYITLSSLINQNGTFFWQGNQTTNTFLSGAGSGAMSSSVSWTADFDGDM 4715

QY 626 DLWLAASSG-----LYHSTNG-----641
DB 4716 DLFLPASQGRANYGSLLENTNGVLGCPVAVGATATTATYASQFSLAVDWNHDLGLMDIARIAQ 4775
```

Qy 471 V-FTTGTSVDYAEINLPSIIVRAGSFDPSQPNDRHVAFTS---DGGKNWFQSGEPG-----522

Db 4670 IQVTGVTGLTWYGSi-----VAFDKBGGDGLYDFWIGDAGGSPDSNT 4708

Qy 523 -----GVTTGGTVAASADGSRFVWAPGDPGVVYAVVFGNSWAASQGVPAQAQLRSR 576

Db 4709 FLNNAGTLVGNSTTNSGGSATVGG-----AVTGYLSLNEGSGVDLNDGRIDL 4758

Qy 577 V-----NPKTFYALS-----NGTFYRSTGGVTFPQVPAAGLSPSSGAVGYMFHAVPGK-EG 625

Db 4759 VQHTYLNLYNTLSSLLNQNGTFFVWQNTTNTFLSGAGSGAMSSSVMTWADFDDGDM 4818

Qy 626 DIWLALASSG-----LYHSTNG-----641

Db 4819 DLFLPASQQRANYGSLLENTNGVLCVPAVAGTATYTSQPSLAVDNNHGLMDIARIAQ 4878

Qy 642 -GSSWGAITGVSSAVNVGFKSAPSSYPVAVFVVGVTIGGVT-----GAYR---SDDCG 690

Db 4879 TGQSY-LYTNVSNASN---WTQSALGGSQS-----GTTSGVAAMDYWDGAVDVLVSKQSG 4930

Qy 691 TTWVLINDDQHQYGNMGQAITGDHANLRRYVIGTNGRGVIVYGD 733

Db 4931 SVFLSRNTNTSYG-----TSLHLRI-----TDPNGINVIVYGN 4963

RESULT 15

A28452

proteoglycan core protein precursor, cartilage - rat

N;Alternate names: aggrecan

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 21-Jan-2000

C;Accession: A92623; A28335; A28453; A28095; A28452

R;Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.

J. Biol. Chem. 262, 17757-17767, 1987

A;Title: Complete primary structure of the rat cartilage proteoglycan core protein

A;Reference number: A92623; MUID:88087070; PMID:3693370

A;Accession: A92623

A;Molecule type: mRNA

A;Residues: 1-2124 <DOE>

R;Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.

J. Biol. Chem. 263, 10040a, 1988

A;Reference number: A30069

A;Contents: annotation; revision to residue 698

R;Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.

J. Biol. Chem. 261, 8108-8111, 1986

A;Title: Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan aggregates. The link protein and proteoglycan core protein

A;Reference number: A28335; MUID:86250698; PMID:2424893

A;Accession: A23835

A;Molecule type: mRNA

A;Residues: 1856-2124 <DO2>

A;Cross-references: GB:M13518; NID:g206104; PIDN:AAA1836.1; PID:g206105

R;Neame, P.J.; Christner, J.E.; Baker, J.R.

J. Biol. Chem. 262, 17768-17778, 1987

A;Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan core protein

A;Reference number: A28453; MUID:88087071; PMID:3693371

A;Accession: A28453

A;Molecule type: protein

A;Residues: 20-37, 'W', 39-60, 'E', 62-64, 'X', 66-69, 70-83, 84, 89-148, 'L', 150-238, 'S', 240-241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778,



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 20.1745 Seconds  
(without alignments)

1909.933 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTYTWSNVAIGGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	168	4.2	2124	1 PGCA RAT	P07897 rattus norv
2	167.5	4.2	1119	1 ALS3 CANAL	O74623 candida alb
3	166	4.1	1468	1 N153 RAT	P49791 rattus norv
4	157	3.9	809	1 QUIA ACICA	Q59086 acinetobact
5	154.5	3.8	5703	1 MUSEB HUMAN	Q3hc84 homo sapien
6	153	3.8	1150	1 APMU PIG	P12021 sus scrofa
7	152	3.8	1196	1 AMYE PAEPO	P21543 paenibacill
8	149	3.7	2132	1 ISOA PSEAY	Q61282 mus musculu
9	148.5	3.7	776	1 ISOA PSEAY	P10342 pseudomonas
10	148.5	3.7	776	1 ISOA PSEAY	P26501 pseudomonas
11	148.5	3.7	1953	1 BIGA SALTY	P25927 salmonella
12	146	3.6	2109	1 PGCA CHICK	P07898 gallus gall
13	145.5	3.6	790	1 QUIA XANCJ	Q9xd78 xanthomonas
14	145	3.6	342	1 Y48L SYNV3	P73069 synechocyst
15	145	3.6	747	1 GUND CELFI	P50400 cellulomona
16	144.5	3.6	708	1 QHED COMTE	Q46444 comamonas t
17	143.5	3.6	937	1 NU98 RAT	P49793 rattus norv
18	142.5	3.5	995	1 AGAA VIBES7	P48839 vibrio sp.
19	142	3.5	647	1 NANH MICVI	Q02834 micromonosp
20	142	3.5	757	1 DBET GLUOX	O05542 gluconobact
21	142	3.5	760	1 YHIL ECOLI	P75780 escherichia
22	139.5	3.5	1045	1 GUNE CELFI	P26225 cellulomona
23	139.5	3.5	2333	1 PGCA CANFA	Q28343 canis famil
24	139	3.4	3063	1 CAIC HUMAN	Q99715 homo sapien
25	138.5	3.4	955	1 YCUG ECOLI	P76017 escherichia
26	138.5	3.4	1526	1 Y4V6 ANASP	Q8yri1 anabaena sp
27	138	3.4	827	1 XANP XANS2	Q60106 xanthomonas
28	138	3.4	1159	1 SOR2 MOUSE	Q9ep95 mus musculu
29	137.5	3.4	872	1 GUXA CELFI	P50401 cellulomona
30	137.5	3.4	1103	1 VG37 BPARI	Q9g0b5 bacterioph
31	137.5	3.4	2415	1 PGCA HUMAN	P16112 homo sapien
32	136.5	3.4	522	1 NU62 HUMAN	P37198 homo sapien
33	136.5	3.4	781	1 NANH_VIBCH	P37060 vibrio chol

34	136.5	3.4	1729	1 NU98 HUMAN	P52948 homo sapien
35	136	3.4	2628	1 HGA2_PORGI	Q51845 porphyromon
36	135.5	3.4	725	1 YA33 SULSO	Q97297 sulfolobus
37	134.5	3.3	2344	1 POLN_RHDV	P27410 rabbit hemo
38	134	3.3	890	1 BCN5_CLOPE	P08696 clostridium
39	134	3.3	1014	1 NANH_CLOSE	P29767 clostridium
40	133.5	3.3	966	1 FIB1_PETWA	P02674 petromyzon
41	132	3.3	886	1 VGP3_EBVAB	Q07284 epstein-bar
42	132	3.3	1433	1 SUBF_BACSU	P16397 bacillus su
43	132	3.3	1541	1 ASX1 HUMAN	Q81xj9 homo sapien
44	131.5	3.3	1026	1 VG37_BPT4	P03744 bacterioph
45	131.5	3.3	1199	1 P121_RAT	P52591 rattus norv

## ALIGNMENTS

RESULT 1  
PGCA RAT ID PGCA RAT STANDARD; PRT; 2124 AA.  
AC P07897;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE AggreCan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).  
GN AGC1 OR AGC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_TaxID=10116;  
RX MEDLINE=8087070; PubMed=3693370;  
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RT "Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones."  
RL J. Biol. Chem. 262:17757-17767(1987).  
RN [2]  
RP REVISION TO 698.  
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RL J. Biol. Chem. 263:10040-10040(1988).  
RN [3]  
RX MEDLINE=86250698; PubMed=242493;  
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;  
RT "Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan."  
RL J. Biol. Chem. 261:8108-8111(1986).  
CC -!- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.  
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).  
CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.  
CC -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate chains. N-linked and O-linked oligosaccharides.  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC -!- SIMILARITY: Contains 4 link domains.  
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



RESULT 2  
 ALS3 CANAL  
 ID ALS3 CANAL STANDARD; PRT; 1119 AA.  
 AC 074623;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Agglutinin-like protein 3 precursor.  
 GN ALS3.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1161.  
 RX MEDLINE=98309840; PubMed=9644209;  
 RA Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;  
 RT "Candida albicans ALS3 and insights into the nature of the ALS gene  
 family";  
 RL Curr. Genet. 33:451-459(1998).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -!- PTM: N-glycosylated and O-glycosylated (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U87956; AAC39486.1; --  
 DR InterPro; IPR008440; Candida ALS.  
 DR Pfam; PF05792; Candida ALS; 1.  
 KW Cell adhesion; Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 1119  
 FT DOMAIN 433 792  
 FT REPEAT 433 468  
 FT REPEAT 469 504  
 FT REPEAT 505 540  
 FT REPEAT 541 576  
 FT REPEAT 577 612  
 FT REPEAT 613 648  
 FT REPEAT 649 684  
 FT REPEAT 685 720  
 FT REPEAT 721 756  
 FT REPEAT 757 792  
 FT DOMAIN 399 404  
 FT DOMAIN 450 455  
 FT DOMAIN 557 563  
 FT DOMAIN 593 597  
 FT DOMAIN 630 635  
 FT DOMAIN 666 671  
 FT DOMAIN 702 707  
 FT DOMAIN 738 743  
 FT DOMAIN 774 777  
 FT DOMAIN 1044 1047  
 FT CARBOHYD 471 471  
 FT CARBOHYD 543 543  
 FT CARBOHYD 579 579  
 FT CARBOHYD 651 651  
 FT CARBOHYD 687 687  
 FT CARBOHYD 723 723  
 FT CARBOHYD 759 759  
 FT CARBOHYD 845 845  
 FT CARBOHYD 987 987  
 FT CARBOHYD 1050 1050  
 FT CARBOHYD 1061 1061  
 FT CARBOHYD 1119 AA; 119927 MW; 6A3FB3FC8C79A71 CRC64;  
 SQ SEQUENCE

Query Match 4.2%; Score 167.5; DB 1; Length 1119;  
 Best Local Similarity 18.7%; Pred. No. 0.03;  
 Matches 170; Conservative 98; Mismatches 340; Indels 303; Gaps 42;  
 QY 7 TMSNVAIGGGFVDCGIVFNEGAPGLYVYRTDIDGG-----MYRWDANGRW--IPL-LDMV 58  
 DB 264 TYKNVPAGYRPFVDAYI-----SATDVNSYTLSEYNEVTCAGGYQWRAPFLRWT 313  
 QY 59 GWNN--WGYNGVVSIAADPINTNKVAAVGMVYTNWDNDGAILRSDQCATWQI--TPLP 115  
 DB 314 GYRNSDAGSNGIVIVATRTVDSITAVT--TLPEDPN-----RDKTKTIELKPIP 363  
 QY 116 -----FKLGNNMPGRGMRERLANDPNDNILYFGAPSGKGLWRSTDSGAT----- 160  
 DB 364 TTTITTSYGVGVTTSYTKTAPIGETATVIV--DIPYHTTTVTSKWTGTTTTHTNP 420  
 QY 161 -----WSQMTNFPDVGIVIANP--TD-----TTG 182  
 DB 421 TDSIDTVIVQVSPNPNTVTTEYWSQ--SFATTTITGPPGNTDTVLIREPPNHTVTTE 478  
 QY 183 YOSDIQGVVWVAFDKSSSSSLGQASKTIFGVADPNPV-----FMSRD--GGATWQAVPG 235  
 DB 479 YWSE-----SYTTSTFTAPPGTDSVIKEPPNPVTTEYWSESYTTSTFTAPPG 531  
 QY 236 APTGFIPIHKGVPDPVNHVL-----YIATSTNGPYDG-----SSGDV 272  
 DB 532 GTDSVI---IKEPPNHTVTTEYWSQSYTTTITVAPPGGDTVLVREPPNHTVTTEY 587  
 QY 273 WKFSVTSCTWRISVPVSTD-----TANDYFYSGLT-----ID 306  
 DB 588 WSQSYTT--TTTVIAPPGGTDVIRREPPNPVTTEYWSQSYATTTITAPPGETDTVLI 646  
 QY 307 ROHPNTINVAQISWV-----PDRIIFRSTDGATWTRIDWTSYENRS 350  
 DB 647 REPPNHTVTTE--YWSQSYATTTITAPPGETDTVLIREPPNHTVTT-----TEWSQS 699  
 QY 351 LRYVLIDISAEPLMTFGV---QPNPPVPSPKLGWMD-----AMADPFNSDRML----- 396  
 DB 700 FATTTTVPAPGGTDTVIRREPPNHTVTTEY--WSQSYATTTITAPPGETDTVLIREPP 758  
 QY 397 -----YGTGATLVATNDLTK-----WDSGQIHIAPMVKG 426  
 DB 759 NHTVTTEYWSQSYATTTITAPPGETDTVLIREPPNPVTVTTEYWSQ----- 806  
 QY 427 LEETA VNDLISPPSCAPLI-----SALDGLGFT-----HADVTAVPSTFTSPVPTTG 475  
 DB 807 -SYTTATTTVPAPGGTDTVIYDTWSSSEISFSRPHYTNHTLTWTTTWTIETKTIETS 865  
 QY 476 TSVD---YAEINPSIIVRAGSFPDSSQPNDRHVPSTDGKNWFOGSPGGVTTGGTVA 531  
 DB 866 CEGDKGCSWVSSTRIVTIPNNIETPMVTN-----TVDSSTTESQSPSGIFSESGVS 919  
 QY 532 ASADGSRFVWAPGDPGQP-----VYVAVFGNSWAASQGVANAAQIRSDRNPKTFFALS 586  
 DB 920 VETESSTVTTAQTNPSVPTTESEVVF-----TTKNNENGPYESPSTNVKSSMD-E 969  
 QY 587 NGTFVRSTDDGVTFOFVAAGLPSSCAVGMFHAVPCKEDLMLAASSGLYHSTNGSSWS 646  
 DB 970 NSEPTTSTAATSTDIENETIATTSV-----EASSPTISSADETTTV 1013  
 QY 647 AITGVSSAV-----NVGFGKSAPOSSYPVAVFVVGITGGVTCAYRSDDCQTTWVLINDQ 700  
 DB 1014 TTTAESTSVIEQPTNNGGKAPSATSPSTTTTANDSVI-----TGTSTNQSQSQ 1066  
 QY 701 HQYGNWGOAIT 711  
 DB 1067 SQYNSDTQOTT 1077  
 RESULT 3  
 N153 RAT  
 ID N153\_RAT STANDARD; PRT; 1468 AA.





CC -/- COFACTOR: PQQ.  
 CC -/- PATHWAY: Quinic acid catabolic pathway; first step. This pathway  
 CC allows growth of bacteria with quinate by its conversion to  
 CC protocatechuate and subsequent metabolism by the beta-ketoadipate  
 CC pathway.  
 CC -/- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -/- INDUCTION: By protocatechuate.  
 CC -/- SIMILARITY: Belongs to the bacterial PQQ dehydrogenase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L05770; AAC37161.1; -;  
 CC PIR; A55547; A55547.  
 CC InterPro; IPR001479; Bac\_PQQ.  
 CC InterPro; IPR002372; Bac\_PQQ\_repeat.  
 CC Pfam; PF01011; PQQ; 4.  
 CC SMART; SM00564; PQQ; 3.  
 CC PROSITE; PS00363; BACTERIAL\_PQQ\_1; FALSE NEG.  
 CC PROSITE; PS00364; BACTERIAL\_PQQ\_2; FALSE NEG.  
 CC Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.  
 CC TRANSMEM 14 34 POTENTIAL.  
 CC TRANSMEM 41 61 POTENTIAL.  
 CC TRANSMEM 68 88 POTENTIAL.  
 CC TRANSMEM 90 110 POTENTIAL.  
 CC TRANSMEM 127 147 POTENTIAL.  
 CC SEQUENCE 809 AA; 88196 MW; 7167CEBEA62BFCB CRC64;  
 CC -----  
 CC Query Match 3.9%; Score 157; DB 1; Length 809;  
 CC Best Local Similarity 20.5%; Pred. No. 0.081;  
 CC Matches 168; Conservative 90; Mismatches 254; Indels 306; Gaps 45;  
 QY 30 GILYVRTDIGGMY-----RWDAANGRIWPLL-----DWGVNNGYN--GVVSTAADP 75  
 DB 131 GGLTVUGMLGGLGMPHPIETVKAASEELPLVPDPAKKVNDVNDYNDAGGRFVALDQ 190  
 QY 76 INTNKVAAVGMVYTNSDPNDPNDGAILRSSDQATWQITPLPFLKGGNMPGRGMGERLAVDP 135  
 DB 191 INENN-----VSKLEAWRFRTGDTTCTGNGAEDQMTPLQ-----VGNKVFLCT 235  
 QY 136 NNDNLYFGAPSGKGLWRS-----TDSGAT----- 160  
 DB 236 PHNNIFAIDADSQGLWKAENVSTADAWERCGRGVAVFDSQPLVQPTLAGATPVAALAN 295  
 QY 161 -----WSQMTNFPDVGTYIANPTDTTGQSD--IQGVVWVAFDKSSSLGQASK----- 207  
 DB 296 TECPRRVYNTVD--GRLLIADVNTGACKDFGVNGTV-----NLHGLGENTKAPREVT 349  
 QY 208 -----TIFVG-----VADPNPNVFWSRD---GGATWQAVPGAPTGFIPHKGVFDP 249  
 DB 350 SAPIAGTIIVGSRADNVAAADMPGVIAYDVITGKLKWAEDPRNP-----DP 399  
 QY 250 VNHVLIATSNTPGGYDSSGDVWKFVSVTGTRISPVSTDT-----ANDYFGYS 301  
 DB 400 -NVVL-----KPGEIVKRSSTN--SWAAMSVDPMQNTVFLPMGSSSVVWVWGN 444  
 QY 302 GLTIDRQHNTINVAQISWPDPTIIFRSTDDGATW-----TRINDWTSYPNRSRYVL 355  
 DB 445 RTAADHKY-NTSVLALD-----ATTGKEKVVYNTVHNDLWDF----- 480  
 QY 356 DISAEPWLT-FGVQPNPPVPSPKLGWMDAEMADDPFNSDRMLYGTGATLVATNDLTWDS 414  
 DB 481 DLPNQSLVDFFPKWDTGTFKFAVIG-----TK----- 507  
 QY 415 GGOIHI-----APMVKGLSE-----TAVNDLISPPSGAPLISALGDLDGGF 454  
 DB 508 SGQFYVLDVGTGKPLTKVIEQPIKVAIDPGEQVSKTPQPRVEMPEQIGNQTLKE-SDMWGA 566

QY 455 THADTVAVP-----STIETSPVFTTGTSDVYA-----ELNPSIIIVRAGSEDPSSQ- 499  
 DB 567 TPFQDLNCRINFKSMRYDGLYAP-----GTDVLSLFPQSLGGMWGS1-----AFDPTHY 618  
 QY 500 -----PNDRHVAFTDGGKQWFGQSEPRGGVTTGCTVAASADGSRFVW 541  
 DB 619 MFVNDMLRLGLWLIQIKQTPED--IKIQANGGEKVNTGM--GAVPMKGT-PYKVKRFRMS 673  
 QY 542 APGDGPGQPVVYAVGFGNSWAASQGVPAQAQIRSDRVPKTFYALNSNGTFYRSTDDGVTF- 600  
 DB 674 ALGIP-----CQKPFPGTMTAIDMKTRQVAVQVPLGTITQDTGPMGIKMG 717  
 QY 601 --QPVAGLPSGAV-----GVMFHVPCKGCDLWLAASSGLYHSTNGSSSNAITGV-S 652  
 DB 718 LKAPI--GMPTIGGPMATQGGGLVFPAA---TQDYILRA-----FNSSNGKELMKARLPVGS 768  
 QY 653 SAVNVGCKSAPGSSYPVAVFVVGTGIGVTGAYRSDDCG 690  
 DB 769 QGTPMSYMSPKTGKQY-----VVVSAG---GARQSPDHG 799  
 RESULT 5  
 MUSB\_HUMAN  
 ID MUSB\_HUMAN STANDARD; PRT; 5703 AA.  
 AC Q9HC84; O00447; O00573; O14985; O15494; O95291; O95451; Q14881;  
 AC Q99552; Q9UE28;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High  
 DE molecular weight salivary mucin MGI) (Sublingual gland mucin).  
 GN MUC5B OR MUC5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-1594 FROM N.A.  
 RA Chen Y., Di Y.P., Wu R.;  
 RT "Molecular cloning of the amino-terminal and 5'-flanking region of the  
 RT human MUC5B mucin gene.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-1325 FROM N.A.  
 RX MEDLINE=98002974; PubMed=9790959;  
 RA Offner G.D., Nunes D.P., Keates A.C., Afchal N.H., Troxler R.F.;  
 RT "The amino-terminal sequence of MUC5B contains conserved  
 RT multifunctional D domains: implications for tissue-specific mucin  
 RT functions.";  
 RL Biochem. Biophys. Res. Commun. 251:350-355(1998).  
 RN [3]  
 RP SEQUENCE OF 40-1324 FROM N.A.  
 RX MEDLINE=98023932; PubMed=9804771;  
 RA Desseyn J.-L., Buissine M.P., Porchet N., Aubert J.-P., Laine A.;  
 RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic  
 RT sequences upstream of the large central exon.";  
 RL J. Biol. Chem. 273:30157-30164(1998).  
 RN [4]  
 RP SEQUENCE OF 1326-4895 FROM N.A.  
 RX TISSUE=Placenta;  
 RA MEDLINE=97166151; PubMed=9013550;  
 RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,  
 RA Laine A.;  
 RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes  
 RT various alternate subdomains resulting in a super-repeat. Structural  
 RT evidence for a lip15.5 gene family.";  
 RL J. Biol. Chem. 272:3168-3178(1997).  
 RN [5]  
 RP SEQUENCE OF 4057-4480 FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=97292540; PubMed=9147051;  
 RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,  
 RA Hannibal J., Clausen H.;

RT "Identification of a major human high molecular weight salivary mucin (MGI) as tracheobronchial mucin MUC5B.";  
 RL Glycobiology 7:413-419(1997).  
 RN [6]  
 RP SEQUENCE OF 4721-5703 FROM N.A.  
 RC TISSUE=Gall bladder;  
 RX MEDLINE=97293229; PubMed=9164870;  
 RA Keates A.C., Nunes D.P., Aidhal N.H., Troxler R.F., Offner G.D.;  
 RT "Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B.";  
 RL Biochem. J. 324:295-303(1997).  
 RN [7]  
 RP SEQUENCE OF 4809-5687 FROM N.A.  
 RC TISSUE=Sublingual gland;  
 RX MEDLINE=96125355; PubMed=8554565;  
 RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;  
 RT "Molecular cloning of a novel high molecular weight mucin (MGI) from human sublingual gland.";  
 RL Biochem. Biophys. Res. Commun. 217:1112-1119(1995).  
 RN [8]  
 RP SEQUENCE OF 4859-5703 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=97347489; PubMed=9201995;  
 RA Desseyn J.-L., Aubert J.-P., Forchet N., Laine A.;  
 RT "Genomic organization of the 3 region of the human MUC5B mucin.";  
 RL J. Biol. Chem. 272:16873-16883(1997).  
 CC -!- FUNCTION: Salivary mucin that is thought to contribute to the lubricating and viscoelastic properties of whole saliva.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also in submaxillary glands, endocervix, gall bladder, and pancreas.  
 CC -!- PTM: Highly glycosylated.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 3 WFC domains.  
 CC -!- SIMILARITY: Contains 4 WFD domains.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements> or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 DR EMBL: AF107890; AAC33673.1; -;  
 DR EMBL: AF086604; AAC87545.1; -;  
 DR EMBL: AJ004862; CAA06167.1; -;  
 DR EMBL: Z72496; CAA96577.1; -;  
 DR EMBL: X74955; CAA52910.1; -;  
 DR EMBL: U63836; AAB61398.1; -;  
 DR EMBL: U78554; AAC51344.1; -;  
 DR EMBL: U78552; AAC51344.1; JOINED.  
 DR EMBL: U78553; AAC51344.1; JOINED.  
 DR EMBL: U78551; AAC51343.1; -;  
 DR EMBL: U95031; AAB65151.1; -;  
 DR EMBL: Y09788; CAA70926.1; -;  
 DR Genew: HGNC:7516; MUC5B.  
 DR MIM: 600770; -;  
 DR GO: GO:0005515; F:protein binding; IPI.  
 DR InterPro: IPR006208; Cys\_knot.  
 DR InterPro: IPR006207; Cys\_knot\_C.  
 DR InterPro: IPR009041; RWE\_inhibitor.  
 DR InterPro: IPR002919; TIL\_Cysrich.  
 DR InterPro: IPR006552; VC\_out.  
 DR InterPro: IPR001007; WVF\_C.  
 DR InterPro: IPR001846; WVF\_D.  
 DR Pfam: PF00007; Cys\_knot; 1.  
 DR Pfam: PF01826; TIL; 1.  
 DR Pfam: PF00093; wvc; 1.  
 DR Pfam: PF00094; wvd; 4.  
 DR SMART: SM00214; WVC; 6.  
 DR SMART: SM00215; WVC\_out; 4.

DR SMART: SM00216; WVD; 4.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; WFC\_1; 2.  
 DR PROSITE: PS0184; WFC\_2; 2.  
 KW Glycoprotein; Repeat; Signal; Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 5703 MUCIN 5B.  
 FT DOMAIN 77 225 WVD 1.  
 FT DOMAIN 329 386 TIL.  
 FT DOMAIN 426 580 WVD 2.  
 FT DOMAIN 858 918 WFC 1.  
 FT DOMAIN 896 1044 WVD 3.  
 FT DOMAIN 1457 1603 THR-RICH.  
 FT DOMAIN 1609 4873 THR-RICH.  
 FT DOMAIN 5005 5178 WVD 4.  
 FT DOMAIN 5353 5425 WFC 2.  
 FT DOMAIN 5482 5528 WFC 3.  
 FT DOMAIN 5594 5683 CTCK.  
 FT DISULFID 5594 5646 BY SIMILARITY.  
 FT DISULFID 5622 5660 BY SIMILARITY.  
 FT DISULFID 5626 5676 BY SIMILARITY.  
 FT DISULFID 5643 5678 BY SIMILARITY.  
 FT DISULFID 5645 5682 BY SIMILARITY.  
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 806 806 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 930 930 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1557 1557 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1775 1775 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3419 3419 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3948 3948 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4745 4745 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4901 4901 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4958 4958 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4965 4965 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4987 4987 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5037 5037 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5052 5052 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5156 5156 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5427 5427 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5467 5467 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5506 5506 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5507 5507 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5543 5543 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5553 5553 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5604 5604 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5618 5618 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5662 5662 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARIANT 5137 T -> S (in dbSNP:2672788).  
 FT CONFLICT 34 G -> E (IN REF. 2).  
 FT CONFLICT 95 PFGLCN -> LPCLCK (IN REF. 2).  
 FT CONFLICT 104 S -> C (IN REF. 2).  
 FT CONFLICT 142 E -> K (IN REF. 1).  
 FT CONFLICT 225 R -> S (IN REF. 2).  
 FT CONFLICT 330 PL -> T (IN REF. 2).  
 FT CONFLICT 337 E -> N (IN REF. 2).  
 FT CONFLICT 356 E -> K (IN REF. 2).  
 FT CONFLICT 362 G -> R (IN REF. 2).  
 FT CONFLICT 369 MISSING (IN REF. 2 AND 3).  
 FT CONFLICT 374 D -> N (IN REF. 2).  
 FT CONFLICT 394 RT -> TR (IN REF. 2).  
 FT CONFLICT 468 RK -> GR (IN REF. 2).  
 FT CONFLICT 512 L -> P (IN REF. 2).  
 FT CONFLICT 585 GAA -> AH (IN REF. 3).

```
FT CONFLICT 601 601 A -> S (IN REF. 3).
FT CONFLICT 629 629 DP -> RS (IN REF. 2).
FT CONFLICT 633 633 F -> L (IN REF. 2).
FT CONFLICT 676 676 A -> P (IN REF. 3).

Query Match 3.8%; Score 154.5; DB 1; Length 5703;
Best Local Similarity 20.2%; Pred. No. 1.2;
Matches 130; Conservative 58; Mismatches 281; Indels 173; Gaps 23;

Qy 102 SSDGATWQL-----TLPFKLGNGMPGRGRLAVDPNNDNLYFCAP 146
Dy |||||
Dy 2956 SSTGTTWLTQTAATTAATTGTTAIPSTPGTAPPKVLTSQATPT-----AT 3007
Qy 147 SGKGLWRST-----DSGATWSQMTNPPDVGTVIANPTDTTGYQSDIQGVVWVAFD 196
Dy |||||
Dy 3008 SSKATSSSPRTATTLPLVTSTATKSTATSTPTIPSSLTGTTGTSQNRPPHPMATMSTIH 3067
Qy 197 KSSSLGQAKTIFV-----GVADPNPVFWRDGGATW-----QAVPGAPTGFPHKG 245
Dy |||||
Dy 3068 PSSTPETHTSTVLTKATTTTRATSSMSTP---SSTPGTTWLTTELTAATTAALPH-- 3122
Qy 246 VFDPNVHLYIATNSGTGPDYSGSDYWKFSVTSGTWTRISPVSTDTANDYFGVGLTI 305
Dy |||||
Dy 3123 -----GTPSSTPGTTWLTPEPSTTATVTPTGSTATASS----- 3156
Qy 306 DRQHPNTIMVATOISWMPDTTIFRSTDGGATWTRIMDWTSPNRSRLRYVLDISAEPMLTF 365
Dy |||||
Dy 3157 TRATAGTLKVLSTATPTPTVLISSRATSSSPGTA---TALP--ALRSTATTPTATSVT- 3209
Qy 366 GVQPNPVPVPGKLG--WDEMAIDFPNSDRMLYGTGATLYATNDLTKWDSGGQIHAPM 423
Dy |||||
Dy 3210 -----AIPSSSLGTAWRLTSQTTP-----TATMSTATPSSTP---ETVHTSTV 3250
Qy 424 VKGLEETAVNDLISPPSGAPLISALDGLGTHADVTAPVS-----TFTSPVFTTGISVD 479
Dy |||||
Dy 3251 LTTATTATRTGVSATPSSTPGTATTKVPTTTTGTATPSSPGTALTTPVWISTTTTP 3310
Qy 480 YAEINPSIIVRAGSFDPSQNDRHVA-----FSTDGKNWFGSGEPGGVT 525
Dy |||||
Dy 3311 -----TTRGSTVTSSIPGTHATVLTTLTTTATGATMPTSSSTQTSSTGTPPSLT 3362
Qy 526 TGGTVAASADGSRFVWAPG-DPGQVYVAVGVGNSWAASQGVPAANAQIRSDRVNPKTFYA 584
Dy |||||
Dy 3363 TATTITATGSTTNFSSPTGTTPIPVLTATTATPAATSSVTTPSSAL----- 3409
Qy 585 LSGTFRSTDDGGVTFQVPAAGLPSGAGVGFHAPVKGEDLWLAASGLYHSTNGSS 644
Dy |||||
Dy 3410 ---GTTHTPPVPNTTATHGRSLPSSP-----HTVP---TAWTSATSGILGTH--- 3453
Qy 645 WSAITGVSSAVNVGFGKSAPSSYPVAVVGTIGGVTVGAYBS 686
Dy |||||
Dy 3454 ---ITEPST-----GTSHTPAATTGTPFSPALSS 3481

RESULT 6
APMU_PIG STANDARD; PRT; 1150 AA.
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apomucin (Mucin core protein) (Fragment).
OS Sus scrofa (Fig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=91236743; PubMed=2033060;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in addition to a highly repetitive,
```

```
RT glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated, identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE=Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A., Johnson W.C. Jr., Hill R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
RL J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC TISSUE=Submaxillary gland;
RX MEDLINE=97248516; PubMed=9092502;
RA Gerken T.A., Owens C.L., Pasumamthy M.;
RT "Determination of the site-specific O-glycosylation pattern of the porcine submaxillary mucin tandem repeat glycopeptide. Model proposed for the polypeptide:galnac transferase peptide binding site.";
RL J. Biol. Chem. 272:9709-9719(1997).
CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL ENVIRONMENT.
CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A MULTIMERIC MUCIN STRUCTURE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81 RESIDUES.
CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER RESIDUES WHICH HAVE GLY AT POSITIONS +2 OR -2 FROM THE GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO ENHANCE GLYCOSYLATION.
CC -1- SIMILARITY: Contains 1 WFVC domain.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
DR EMBL; M61883; AAA30998.1; --
DR EMBL; M21174; AAA30990.1; --
DR InterPro; IPR006208; Cys knot.
DR InterPro; IPR006207; Cys knot_C.
DR InterPro; IPR01007; WFVC.
DR Pfam; PF00007; Cys knot; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; WFVC_1; 1.
DR PROSITE; PS0184; WFVC_2; 1.
DR Glycoproteins; Repeat.
KW NON_TER 1
FT DOMAIN <1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1.
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
```

```

FT REPEAT 288 368 5. (INCOMPLETE) .
FT REPEAT 369 391 6. (INCOMPLETE) .
FT DOMAIN 929 995 WMFC.
FT DOMAIN 1062 1146 BY SIMILARITY.
FT DISULFID 1062 1109 BY SIMILARITY.
FT DISULFID 1076 1123 BY SIMILARITY.
FT DISULFID 1085 1139 BY SIMILARITY.
FT DISULFID 1089 1141 BY SIMILARITY.
FT DISULFID ? 1145 BY SIMILARITY.
FT CARBOHYD 46 46 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 50 50 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 51 51 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 57 57 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 58 58 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 61 61 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 66 66 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 67 67 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 73 73 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 74 74 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 76 76 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 77 77 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 81 81 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 83 83 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 87 87 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 91 91 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 93 93 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 94 94 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 96 96 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 98 98 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 101 101 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 103 103 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 104 104 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 106 106 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 107 107 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 108 108 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 110 110 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 114 114 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 117 117 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 123 123 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 124 124 O-LINKED (GALNAC. .) (MUCIN TYPE) .
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 917 917 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. .) (POTENTIAL) .
SQ SEQUENCE 1150 AA; 109615 MW; 3CB68B5D29DD7F5A CRC64;

Query Match 3.8%; Score 153; DB 1; Length 1150;
Best Local Similarity 21.7%; Pred. No. 0.21;
Matches 147; Conservative 75; Mismatches 261; Indels 194; Gaps 31;

QY 157 SGATWSQNTNPDVGTGTYANTDTTGYQSDTQGVVWVAFDKS-----SSSLGQASKTIFVG 212
DB 182 SGSTGSSGSPGATGASIGQPTSRISVAGSGAPVSSGASQAAGTSGAGCGTGTASSVG 241
QY 213 VADPNPNVFWRDGGATQWQVPGA-----PTGFPHKGVDFPVNHLVLIATSN 260
DB 242 VTETARP---SVAGSGTTGTGVSAGSGTSSGSGPGATG----ASIGQPTSRISVAGS- 293
QY 261 TGGPYDGGSDGVKFSVTGTRISPVPTSDTANDYFGYGLTIDRQHPNTIMVATQIS 320
DB 294 SGAPVSSGASQAAGTSGAGCGTGTASSVGVTTETARPSVAGSGTT-----GTVSGASGT 347
QY 321 WNPDTIIFRSTDG--GATWTIRINDWTSYPNRSLRVLDISABPWLTFGVQNPVPPSKL 378
DB 348 -----GSSGSPGATGASI---GQPTSRISVAGSGAPVSSGA-----SQAA 388
QY 379 GWMDEAMAIQFNSDRMLYGGTATLY-ATNDLTK-----WDSGGQIHAP 422
DB 389 GTSEATTSIGAGTSGGVGFKTEATTFPGENETIRVGIATGTTGIVSRKTLBPGSYNTBAT 448
QY 423 MVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSP-----VFTTGT 477

```

```

Db 449 TSIGRSGTHHDL---PGGTTIV-----LPGFHSQSQSSKPGSVTTTPGSPESGETGTS 500
QY 478 VDYA-----ELNPSIIVRAGSFDPSPQNDRHV 505
Db 501 GFSTVITVSGSHTTATTFIGSGSPGTCGSRPCTTGELSGTTASGNATTEATSTETRI 560
QY 506 AFSTDGKKNWFOGSEPGGVTTG-----GTVAASAD-----GSRFVWAPG 544
Db 561 GPOT-GAQTVPQSVQSGSETGTSEAVSNPAIASGSSSTGTTSGASDSQVTGSR---TG 615
QY 545 DPGQPVVAVGFG-NSWAASQGVPAQAQIRSDRVNPKTFYALNNGTFRSTDDGVTQPV 603
Db 616 TTGVVLGTTVAPCSSTGATTGVLNEGTRSTSLGTT---RVASGTTYES-----GT 664
QY 604 AAGLPSGAGVGVMFHAPVKGEDGLWLAASSGLYHSTNGSSWSAITGVSSAVN-VGFPGKS 662
Db 665 SNSVPSGGS-----GTFG-----SGI-----NTGSSQTQVGIQTGTTAVGFGST 704
QY 663 -APGSSYPAVF-----VVGITGGVGTAYRSDCGTTWVLINDDQHQYGNWG 707
Db 705 LLPGSSNTGATTSPBERTSPGSKTGTITRVVSGTTVAGSSSNTGATTSLGRGETTQ-GGIK 763
QY 708 QAITGDHANLRVYIGT 724
Db 764 IVITG-----VTGVT 773

RESULT 7
AMTB_PAEPO
ID AMTB_PAEPO STANDARD; PRT; 1196 AA.
AC P21543;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Beta/alpha-amylase precursor [includes: Beta-amylase (EC 3.2.1.2);
OS Paenibacillus polymyxa (Bacillus polymyxa)].
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE OF 1-936 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=72;
RX MEDLINE=87165765; PubMed=2435707;
RA Kawazu T., Nakanishi Y., Uozumi N., Sasaki T., Yamagata H.,
TSukagoshi N., Udaoka S.;
RT "Cloning and nucleotide sequence of the gene coding for enzymatically
active fragments of the Bacillus polymyxa beta-amylase.";
RL J. Bacteriol. 169:1564-1570(1987).
RN [2]
RP SEQUENCE OF 689-1196 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=72;
RX MEDLINE=89123046; PubMed=2464578;
RA Uozumi N., Sakurai K., Sasaki T., Takekawa S., Yamagata H.,
TSukagoshi N., Udaoka S.;
RT "A single gene directs synthesis of a precursor protein with beta-
and alpha-amylase activities in Bacillus polymyxa.";
RL J. Bacteriol. 171:375-382(1989).
RN [3]
RP SEQUENCE OF 1-776 FROM N.A.
RC STRAIN=ATCC 8523;
RX MEDLINE=87231094; PubMed=2438660;
RA Rhodes C., Strasser J., Friedberg P.;
RT "Sequence of an active fragment of B. polymyxa beta amylase.";
RL Nucleic Acids Res. 15:3934-3934(1987).
RN [4]
RP DISULFIDE BOND, AND MUTAGENESIS OF CYSTEINE RESIDUES.
RX MEDLINE=91215008; PubMed=1827035;
RA Uozumi N., Matsuda T., Tsukagoshi N., Udaoka S.;
RT "Structural and functional roles of cysteine residues of Bacillus
polymyxa beta-amylase.";
RL Biochemistry 30:4594-4599(1991).
CC -!- FUNCTION: THE PRECURSOR PROTEIN IS PROTEOLYTICALLY CLEAVED TO

```

CC	PRODUCE MULTIFORM BETA-AMYLASES AND A 48 kDa ALPHA-AMYLASE AFTER SECRETION.	
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in polysaccharides so as to remove successive maltose units from the non-reducing ends of the chains.	
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.	
CC	-!- SUBCELLULAR LOCATION: Secreted.	
CC	-!- SIMILARITY: In the N-terminal section; belongs to family 14 of glycosyl hydrolases.	
CC	-!- SIMILARITY: In the C-terminal section; belongs to family 13 of glycosyl hydrolases.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
CC	EMBL; M15817; AAA85446.1; -;	
CC	EMBL; Y00150; CAA68344.1; -;	
CC	PIR; A29130; A29130.	
CC	HSSP; P36924; 1B9Z.	
CC	InterPro; IPR006589; Alp_aml_cat_sub.	
CC	InterPro; IPR006048; Alpha_aml_C.	
CC	InterPro; IPR006047; Alpha_aml_cat.	
CC	InterPro; IPR005085; CBM_25.	
CC	InterPro; IPR006046; Glyco_hydro_13.	
CC	InterPro; IPR001554; Glyco_hydro_14.	
CC	Pfam; PF00128; alpha-amylase; 1.	
CC	Pfam; PF02806; alpha-amylase_C; 1.	
CC	Pfam; PF03423; CBM_25; 2	
CC	Pfam; PF01373; Glyco_hydro_14; 1.	
CC	PRINTS; PR00110; ALPHAAMYLASE.	
CC	PRINTS; PR00750; BETAAMYLASE.	
CC	SMART; SM00642; Aamy; 1.	
CC	SMART; SM00632; Aamy_C; 1.	
CC	PROSITE; PS00506; BETA_AMYLASE_1; 1.	
CC	PROSITE; PS00679; BETA_AMYLASE_2; 1.	
KW	Multifunctional enzyme; Hydrolase; Glycosidase; Signal;	
KW	Polysaccharide degradation; Repeat.	
FT	SIGNAL 1 35	
FT	CHAIN 36 1196	BETA/ALPHA-AMYLASE.
FT	DOMAIN 36 454	BETA-AMYLASE.
FT	REPEAT 455 558	
FT	REPEAT 565 668	
FT	DOMAIN 669 1196	ALPHA-AMYLASE.
FT	DISULFID 118 126	
FT	ACT_SITE 198 198	BY SIMILARITY.
FT	ACT_SITE 394 394	BY SIMILARITY.
FT	MUTAGEN 118 118	C->S: 5-FOLD DECREASE IN ACTIVITY.
FT	MUTAGEN 126 126	C->V: 20-FOLD DECREASE IN ACTIVITY.
FT	MUTAGEN 358 358	C->S: 60-FOLD DECREASE IN ACTIVITY.
FT	CONFLICT 1 1	M -> MIGL (IN REF. 3).
FT	CONFLICT 67 67	N -> S (IN REF. 3).
FT	CONFLICT 100 100	N -> D (IN REF. 3).
FT	CONFLICT 154 154	S -> N (IN REF. 3).
FT	CONFLICT 177 177	E -> Q (IN REF. 3).
FT	CONFLICT 227 228	NA -> KS (IN REF. 3).
FT	CONFLICT 330 330	G -> S (IN REF. 3).
FT	CONFLICT 425 425	N -> S (IN REF. 3).
FT	CONFLICT 493 493	D -> A (IN REF. 3).
FT	CONFLICT 532 532	S -> L (IN REF. 3).
FT	CONFLICT 559 559	A -> T (IN REF. 3).
FT	CONFLICT 665 665	A -> T (IN REF. 3).
FT	CONFLICT 681 681	D -> N (IN REF. 3).
FT	CONFLICT 686 686	T -> A (IN REF. 3).
FT	CONFLICT 725 728	AFTS -> VFSP (IN REF. 3).
FT	CONFLICT 736 736	N -> K (IN REF. 3).
FT	CONFLICT 741 741	N -> S (IN REF. 3).
FT	CONFLICT 758 758	S -> N (IN REF. 3).

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
ON NCBI\_TaxID=10090;  
RX STRAIN=FROM N.A.  
RP STRAIN=BALB/c; TISSUE=Cartilage;  
RC MEDLINE=95104847; PubMed=7806222;  
RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fuellep C., Horvath P.,  
RA Dooge K.J., Glant T.T.;  
RT "Complete coding sequence, deduced primary structure, chromosomal  
RT localization, and structural analysis of murine aggrecan.";  
RL Genomics 22:364-371(1994).  
RN [2]  
RP SEQUENCE OF 211-326 FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=95004579; PubMed=7920633;  
RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A.,  
RA Yamada Y.;  
RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in  
RT the aggrecan gene.";  
RL Nat. Genet. 7:154-157(1994).  
RN [3]  
RP INTERACTION WITH FBLN1  
RX MEDLINE=99329059; PubMed=10400671;  
RA Asberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;  
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and  
RT versican.";  
RL J. Biol. Chem. 274:20444-20449(1999).  
CC -!- FUNCTION: This proteoglycan is a major component of extracellular  
CC matrix of cartilaginous tissues. A major function of this protein  
CC is to resist compression in cartilage. It binds avidly to  
CC hyaluronic acid via an amino-terminal globular region. May play a  
CC regulatory role in the matrix assembly of the cartilage.  
CC -!- SUBUNIT: Interacts with FBLN1.  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By  
CC similarity).  
CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino  
CC terminus of the proteoglycan, while another globular region, G3,  
CC makes up the COOH terminus. G1 contains link domains and thus  
CC consists of three disulfide-bonded loop structures designated as  
CC the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)  
CC and the chondroitin sulfate (CS) attachment domains lie between G2  
CC and G3.  
CC -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate  
CC chains. N-linked and O-linked oligosaccharides.  
CC -!- DISEASE: Defects in AGC1 are the cause of cartilage matrix  
CC deficiency (CMD). CMD is an autosomal recessive syndrome  
CC characterized by cleft palate, short limbs, tail and snout.  
CC Mutation in strain CMD causes absence of aggrecan by truncation of  
CC the protein (mutation in the G1 domain).  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC -!- SIMILARITY: Contains 4 link domains.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L07049; AAC37670.1; -;  
DR EMBL; S73722; AB32160.1; -;  
DR EMBL; S73721; AB32160.1; JOINED.  
DR PIR; A55182; A55182.  
DR HSPF; P98066; ITSG.  
DR MGD; MGI:99602; Agc1.  
DR InterPro; IPR002353; AntifreezeII.

DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR InterPro; IPR001304; Lectin\_C.  
DR InterPro; IPR000538; Link.  
DR InterPro; IPR003324; SGXSG.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF02339; SGXSG; 60.  
DR Pfam; PF00084; sushi; 1.  
DR Pfam; PF00193; Xlink; 4.  
DR PRINTS; PR00356; ANTIFREEZEII.  
DR PRINTS; PR01265; LINKMODULE.  
DR ProDom; PD000918; Link; 4.  
DR SMART; SM00032; CCP; 1.  
DR SMART; SM00034; CLECT; 1.  
DR SMART; SM00406; IGV; 1.  
DR SMART; SM00445; LINK; 4.  
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.  
DR PROSITE; PS00835; IG-LIKE; 1.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PROSITE; PS01241; LINK; 4.  
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;  
KW Immunoglobulin domain.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2132 AGGECAN CORE PROTEIN.  
FT DOMAIN 34 147 IG-LIKE V-TYPE.  
FT DOMAIN 170 247 LINK 1.  
FT DOMAIN 268 349 LINK 2.  
FT DOMAIN 504 581 LINK 3.  
FT DOMAIN 602 683 LINK 4.  
FT DOMAIN 1918 2044 C-TYPE LECTIN.  
FT DOMAIN 2048 2106 SUSHI.  
FT DOMAIN 48 140 GI-A.  
FT DOMAIN 152 247 GI-B.  
FT DOMAIN 253 349 GI-B'.  
FT DOMAIN 486 580 G2-B.  
FT DOMAIN 587 682 G2-B'.  
FT DOMAIN 685 803 KS.  
FT DOMAIN 805 1231 CS-1.  
FT DOMAIN 1232 1917 CS-2.  
FT DOMAIN 1917 2132 G3.  
FT DISULFID 51 133 BY SIMILARITY.  
FT DISULFID 175 246 BY SIMILARITY.  
FT DISULFID 199 220 BY SIMILARITY.  
FT DISULFID 273 348 BY SIMILARITY.  
FT DISULFID 297 318 BY SIMILARITY.  
FT DISULFID 509 580 BY SIMILARITY.  
FT DISULFID 533 554 BY SIMILARITY.  
FT DISULFID 607 682 BY SIMILARITY.  
FT DISULFID 631 652 BY SIMILARITY.  
FT DISULFID 1922 1933 BY SIMILARITY.  
FT DISULFID 1950 2042 BY SIMILARITY.  
FT DISULFID 2018 2034 BY SIMILARITY.  
FT DISULFID 2049 2092 BY SIMILARITY.  
FT DISULFID 2078 2105 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT SITE 1171 1173 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDFC6CDBA163 CRC64;

Query Match 3.7%; Score 149; DB 1; Length 2132;  
Best Local Similarity 19.5%; Pred.No. 0.74;  
Matches 163; Conservative 96; Mismatches 286; Indels 292; Gaps 39;

QY 64 GYGVVSTAAADPINTNKVAAVGMVTSWDPNDGAILRSSDQGA-----TWQITPL 114



Db 1105 GYVSGIPSGDGTET-----SASGVEDVSGLPSCGEGLETSASGVEDVSGPSTRDLSAS 1160  
Qy 115 PFKLGNNPGRGMRGLAVDPNDNIIYFGAPSGKGLWRSDTSQATWSQMTNFPD----- 169  
Db 1161 GVDVTGPPSGRGPETSVSGVGD-----PSGLPSKEGLETSSASGA--EDLSGLPSGKEDL 1215  
Qy 170 -----VGYPIANPT-DTTGYQSDIQGVVMAVAFKSSSLGQASKTIFVGVGA 214  
Db 1216 VGSASGALDFKLPPLPGTLGSGQTPVNGPSPGSG-----EYSGADIGSPSS--GLP 1266  
Qy 215 DPNNPVWSRDGATWQAVCAPGTPPHKGVFPVNVHLYIATNTGGPDG----- 267  
Db 1267 D-----FSGLPSPGF-PTVSLVD--STLVEVITATTSSLEGRGTIGIS 1306  
Qy 268 SSGDVWKFVTSQWTRTISVPSTDTANDYFGYSLGTLIDRQHPNTIMVATQISMPWTII 327  
Db 1307 GSGEV-----SGL-----PLGELDSD--ISGL-----PSGTSLGQASGSDS-- 1344  
Qy 328 FRSTDGATWTRIDMTSYENRSLRYVLDISAEPWLTFGVQVPPVSPKLGWMDAMAI 387  
Db 1345 SGETSG-----FFDVSGQFPFGSSGV-----EETSGI 1371  
Qy 388 DPFNSRMLYGTATLYATNDLTKWD--SGQIHIAPMVKGL-----EET 430  
Db 1372 PEISGQPS--GTPDPTT-ATSCVTNELNLSQPDVSGDGSGLFGSGQSGITSVSGETS 1428  
Qy 431 AVNDLIPSPGAPL-----ISALDGLGFTHADYVAPSTIFT----- 468  
Db 1429 GISDLGQSPGFPVFGTATRTPLASGTISGSGESGITFVDTSFVEVPTTFRBEGL 1488  
Qy 469 -----SPVFTGTSDVYAEINPSIIVRAGSPDPSSQPNDRHVAPSTDGKNWF 516  
Db 1489 GSVELSGFPSEGTSLGTSQTVDSQSSGALDSGLTSPTPE-----F 1532  
Qy 517 QGSEPG-----GVTTGCTVAASA-DGSRFWVAPGDP-----GQPVVY 552  
Db 1533 SGLPSGVAEVSSESGVETGSSLPAGFDSGLV--SGFPTVSLVDRTLVESITQAPTAQ 1590  
Qy 553 AVGCGNSW-----AASQGV-----ANQIRSDRVNPKTFALNCGTFYR 592  
Db 1591 EAGEGSGILEFSGAHSGTDPDISGELSGDLSTLQSGQMETSTETSPSPY--FSGDPSS 1648  
Qy 593 STD--GGVTFQPVAAGLPSSGAVGMFHAVPGKEDLMLAASSGLVH----- 637  
Db 1649 TTDVSGESIAATGSGESSGLPEVTLNISELVEGTEPTVSOELGHGSPMYISRLSEAS 1708  
Qy 638 --STNGSSWSAIT-----GVSSAV-----NVGFGKSAFGSSYPVAVFVVGITGG 679  
Db 1709 GDASASGDLGAVTNFPGSGVEASVPEASSDLSAYPEAGVGVSAPEASSKLSFPDLHG 1768  
Qy 680 VTGAYRSDCGTT-----WVLINDQHQVGNWQQAITGDHANLRRVIGTNG 726  
Db 1769 ITSAFHETDLEMTTTPSTEVNSNPWF--QEGTREGSAPEVSGESSTSDIDTGTSG 1823

## RESULT 9

ISOA\_PSEAY STANDARD; PRY; 776 AA.  
AC P10342;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Isoamylase precursor (EC 3.2.1.68).  
GN IAM.  
OS Pseudomonas amylocladum.  
OC Bacteria; Proteobacteria.  
OX NCBI\_TaxID=32043;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB-15;  
RX MEDLINE=88243808; PubMed=3379068;  
RA Amemura A., Chakraborty R., Fujita M., Nouni T., Futai M.;

"Cloning and nucleotide sequence of the isoamylase gene from Pseudomonas amylocladum SB-15.";  
J. Biol. Chem. 263:9271-9275(1988).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JD210;  
RX MEDLINE=91064385; PubMed=2248978;  
RA Chen J.H., Chen Z.Y., Chow T.Y., Chen J.C., Tan S.T., Hsu W.H.;  
"Nucleotide sequence and expression of the isoamylase gene from an isoamylase-hyperproducing mutant, Pseudomonas amylocladum JD210.";  
Biochim. Biophys. Acta 1087:309-315(1990).  
[3]  
RP SEQUENCE OF 744-776 FROM N.A.  
RC STRAIN=SB-15;  
RX MEDLINE=89327147; PubMed=2753857;  
RA Amemura A., Fujita M., Futai M.;  
"Transcription of the isoamylase gene (iam) in Pseudomonas amylocladum SB-15.";  
J. Bacteriol. 171:4320-4325(1989).  
[4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=98387895; PubMed=9719642;  
RA Katsuya Y., Mezaki Y., Kubota M., Matsura Y.;  
"Three-dimensional structure of Pseudomonas isoamylase at 2.2-A resolution.";  
J. Mol. Biol. 281:885-897(1998).  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch linkages in glycogen, amylopectin and their beta-limits dextrans.  
CC -I- COPACITOR: Binds 1 calcium ion per subunit.  
CC -I- SUBUNIT: Monomer.  
CC -I- INDUCTION: By maltose.  
CC -I- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J03871; AAA25854.1; -;  
DR EMBL; X13378; CAA31754.1; -;  
DR PDB; 1BF2; 12-AUG-98.  
DR InterPro; IPR006047; Alpha\_amil\_cat.  
DR InterPro; IPR004193; Glyco\_hydro\_13N.  
DR InterPro; IPR007110; Ig-like.  
DR Pfam; PF00128; alpha-amyase; 1.  
DR Pfam; PF02922; isoamylase N; 1.  
KW Hydrolase; Glycosidase; Calcium-binding; Signal; 3D-structure.  
FT SIGNAL 1 26  
FT CHAIN 27 776  
FT ACT SITE 401 401  
FT ACT SITE 461 461  
FT ACT SITE 536 536  
FT METAL 154 154  
FT METAL 255 255  
FT METAL 256 256  
FT METAL 258 258  
FT METAL 285 285  
FT DISULFID 410 422  
FT DISULFID 546 616  
FT DISULFID 738 766  
FT CONFLICT 8 8  
FT CONFLICT 126 126  
FT CONFLICT 169 171  
FT CONFLICT 386 386  
FT CONFLICT 413 416  
FT CONFLICT 454 489  
A -> G (IN REF. 1).  
F -> C (IN REF. 1).  
GAS -> AH (IN REF. 1).  
L -> V (IN REF. 1).  
GAYT -> AVH (IN REF. 1).  
SGLDLFAEPWAGTSGYQLGGPQGWSEWNLPRDS ->  
TVICILNLGPPSAATRTISWDSRRVVRVSVPRQ (IN  
REF. 1).  
WP -> S (IN REF. 1).  
AFKHAHA -> RSARHIP (IN REF. 1).  
CONFLICT 555 556  
CONFLICT 650 657



FT	TURN	30	31	FT	HELIX	437	444
FT	STRAND	34	37	FT	STRAND	447	447
FT	TURN	39	40	FT	TURN	450	451
FT	STRAND	43	49	FT	STRAND	456	460
FT	STRAND	55	61	FT	TURN	467	468
FT	STRAND	71	74	FT	TURN	472	473
FT	STRAND	76	77	FT	TURN	477	478
FT	TURN	80	81	FT	STRAND	480	482
FT	STRAND	82	88	FT	HELIX	484	495
FT	HELIX	89	94	FT	STRAND	497	497
FT	TURN	95	96	FT	TURN	498	499
FT	STRAND	102	109	FT	STRAND	500	500
FT	TURN	110	111	FT	HELIX	504	511
FT	STRAND	112	112	FT	TURN	512	513
FT	TURN	116	117	FT	HELIX	515	518
FT	TURN	120	121	FT	HELIX	519	521
FT	TURN	124	125	FT	TURN	522	522
FT	TURN	132	133	FT	HELIX	525	527
FT	TURN	139	140	FT	STRAND	528	530
FT	STRAND	143	143	FT	HELIX	540	543
FT	TURN	146	147	FT	TURN	544	545
FT	STRAND	151	151	FT	TURN	555	556
FT	HELIX	163	166	FT	TURN	572	577
FT	STRAND	167	167	FT	HELIX	581	597
FT	TURN	170	174	FT	STRAND	601	605
FT	TURN	178	180	FT	TURN	606	607
FT	STRAND	184	185	FT	HELIX	608	610
FT	STRAND	202	204	FT	TURN	611	611
FT	STRAND	207	210	FT	TURN	615	616
FT	HELIX	212	216	FT	TURN	621	622
FT	TURN	217	218	FT	TURN	626	628
FT	TURN	220	221	FT	HELIX	636	654
FT	HELIX	224	226	FT	HELIX	656	658
FT	TURN	227	228	FT	TURN	666	668
FT	HELIX	230	235	FT	STRAND	669	672
FT	TURN	236	236	FT	TURN	674	675
FT	HELIX	237	243	FT	STRAND	678	678
FT	TURN	244	244	FT	HELIX	681	684
FT	STRAND	247	250	FT	TURN	685	685
FT	STRAND	254	254				
FT	STRAND	256	256				
FT	TURN	259	262				
FT	TURN	265	266				
FT	TURN	269	270				
FT	STRAND	278	278				
FT	STRAND	284	284				
FT	HELIX	286	288				
FT	TURN	294	295				
FT	HELIX	296	310				
FT	TURN	311	312				
FT	STRAND	314	319				
FT	TURN	325	326				
FT	STRAND	329	329				
FT	STRAND	338	339				
FT	HELIX	343	351				
FT	STRAND	352	352				
FT	STRAND	354	354				
FT	TURN	356	357				
FT	STRAND	361	361				
FT	STRAND	371	371				
FT	TURN	373	374				
FT	HELIX	376	391				
FT	TURN	392	392				
FT	STRAND	397	400				
FT	TURN	401	402				
FT	HELIX	403	407				
FT	STRAND	408	408				
FT	TURN	417	418				
FT	TURN	420	421				
FT	TURN	423	424				
FT	STRAND	429	429				
FT	TURN	431	432				
FT	TURN	434	435				

Query Match 3.7%; Score 148.5; DB 1; Length 776;  
Best Local Similarity 20.7%; Pred. No. 0.24;  
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

QY	27	GAPGILYVRTDIGMYRWDAANGRWIPLLDWGNNWNGVVSIAADPINTKVAAVG	86
Db	96	GITGAVY-----YGRWANGPNWPYASNWKGSGQAGFVSDVDANGDRFNPKLLDDPY	147
QY	87	MYTNSWDPNDGAILRSSDQCATWQITPLPFKLGNNPGRGCMGERLAVDPNNNDILYFQAP	146
Db	148	AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP	182
QY	147	SKGLWRSTDSGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS	201
Db	183	KGVWLVPSQTSGTKPTRAQKDDVIYEVHVRGTEQDTSIPAQYRGTYTYGAGLKASYLAS	242
QY	202	LCQASKTIFVGAD-----PN---NPVFW-----SRDGGATWQAVPCAPTG-F	240
Db	243	LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGYMTENYFSPDRRYAYNKAAGGPTAEF	301
QY	241	IPHKGVDFPNHVLVIATSNCTGGPYDGSQGVWKFVSTGTWTRISFVPST-----	291
Db	302	QMVQAFNAGIKVYMDV-----VYNHTAEGGTWTSSTPTTATISWRGLDN	348
QY	292	-----DTANDYF-GYSGLTIDRQHNTI---MVATQISWHPDTIIFRSTGGATWTRIW	341
Db	349	ATYYELTSGNQYFYDNTGIGANFNTYNTVAQNLIIVDSLAWYANTM---GYDGRF----	400
QY	342	DWTS-YPNRSLRYVLDISAEPWLTFCVQPNPVPSPKLGWMDAID-----PNSDRML	396
Db	401	DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDAADSNVAINRILREFTVRPAA	451
QY	397	YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLSE-----TAVNDLISPPSGAPLI	445

```
Db 452 GSGGLDLFA---EPWAIGNSYQLGFPQGWSEWNGLFRDSLROAQNEL-----GSMTI 502
Qy 446 SALGDLGGTHADVAVPSTFTSPVFTTGTSDVYAEINPSIIIV-----AGS 493
Db 503 YVTQDANDFGS-----SNLFQSGSRPWNINFDVHGMTLKOVYSCNGANNSQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPEGGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWDOQMSAG---TGAADVQDQRAARTGMAPEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAQAIRSDR-----V 577
Db 602 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFQAQLIAFRKAHPAL 658
Qy 578 NPKTFYALNSGTFYRSTDDGVTGPVAAAGLPSSGAVG-----VNFHVAVPKGE 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAVAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSVFTFLPAPSGTQWYRVTTDCDWDGASTFV-----APG 751
Qy 666 SSYPAVFVVGTTGGVGTGAYRSDDCGTTWVLINDQHQYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YQCGQSL 770

RESULT 10
ISOA PSESP
ID ISOA PSESP STANDARD; PRT; 776 AA.
AC P26501;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Isoamylase precursor (EC 3.2.1.68).
GN IAM.
OS Pseudomonas sp. (strain SMP1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 27-44.
RX MEDLINE=89381677; PubMed=2778432;
RA Tognoni A., Carrera P., Galli G., Luccchese G., Camerini B.,
RA Grandi G.;
RT "Cloning and nucleotide sequence of the isoamylase gene from a strain
of Pseudomonas sp.;"
RL J. Gen. Microbiol. 135:37-45(1989).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
linkages in glycogen, amylopectin and their beta-limits dextrins.
CC -1- COPACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- INDUCTION: By maltose.
CC -1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25247; AAA25855.1; -
CC EMBL; A10909; CAA00929.1; -
CC HSPB; P10342; IBF2.
CC InterPro; IPR006047; Alpha amyl cat.
CC InterPro; IPR004193; Glyco_hydro_13N.
CC InterPro; IPR007110; IG-like.
CC Pfam; PF00128; alpha-amylase; 1.
CC Pfam; PF02922; isoamylase N; 1.
CC Hydrolase; Glycosidase; Calcium-binding; Signal.
CC SIGNAL
1 26
```

```
FT CHAIN 27 776 ISOAMYLASE.
FT ACT SITE 401 401 BY SIMILARITY.
FT ACT SITE 461 461 BY SIMILARITY.
FT ACT SITE 536 536 BY SIMILARITY.
FT METAL 154 154 CALCIUM (BY SIMILARITY).
FT METAL 255 255 CALCIUM (BY SIMILARITY).
FT METAL 256 256 CALCIUM (VIA CARBONYL OXYGEN) (BY
SIMILARITY).
FT METAL 258 258 CALCIUM (BY SIMILARITY).
FT METAL 285 285 CALCIUM (BY SIMILARITY).
FT DISULFID 410 422 BY SIMILARITY.
FT DISULFID 546 616 BY SIMILARITY.
FT DISULFID 738 766 BY SIMILARITY.
SQ SEQUENCE 776 AA; 83656 MW; A5E4C02EF026A3A4 CRC64;

Query Match 3.7%; Score 148.5; DB 1; Length 776;
Best Local Similarity 20.7%; Pred. No. 0.24;
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

Qy 27 GAPGILYVRTDIGMYRMDAANGRIPLLDVGMNMGYNGVVSIAADPINTNKVMAAVG 86
Db 96 GITGAVY-----YGYRANGPNWPYASNMGKSGQAGFVSDVDANGDRFNENKLLDPY 147
Qy 87 MYTNSWDPNDGAILRSSDQAGATQITPLPKLGGNMPGRGGERLAVDPNNDNLLYFGAP 146
Db 148 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
Qy 147 SGKGLWRSTDSGATWSQMTNFPDV---GTIANPTD--TTGYQSDIOGVVWVAFDKSS--SS 201
Db 183 KGVVLVPSTQSTGTKPTRAQKDDVIYEVHVRGFTQEDTSIPAQYRGTYVAGLKAASYLAS 242
Qy 202 LGQASKTIFVGVD-----PN---NPVFW-----SRDGGATQWQVPGAPTG-F 240
Db 243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMWYENYFSPDRRYAYNKAGGTAEF 301
Qy 241 IPHKGFDPVNVHLYIATSTGGPYDGGSDGVWKFSTVSTGTWTRISPVST----- 291
Db 302 QAMVQAQFHAGIKVYMDV-----VYNHTAEGGTWTSDDPTTIATYSWRGLDN 348
Qy 292 -----DTANDYF-GYSGLTITDRQHPNTI---MVATQISWVWPDITLIFRSTDGATWTRIW 341
Db 349 TTYTYELTSGNQYFYDNTGIGANFNFTNTVAQNLIIVDSLAYWANTM---GVDGFR-----F 400
Qy 342 DWTS-YENRSLRYVLDISAEPWLTFGVQPNPVPSPKLGWMDAEMAD----PFNSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVAINRILREFTVRPAA 451
Qy 397 YGTGATLYATNDLTKWDSGGQTH-IAPMVYKGLEE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA---EPWAIGNSYQLGFPQGWSEWNGLFRDSLROAQNEL-----GSMTI 502
Qy 446 SALGDLGGTHADVAVPSTFTSPVFTTGTSDVYAEINPSIIIV-----AGS 493
Db 503 YVTQDANDFGS-----SNLFQSGSRPWNINFDVHGMTLKOVYSCNGANNSQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPEGGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWDOQMSAG---TGAADVQDQRAARTGMAPEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAQAIRSDR-----V 577
Db 602 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFQAQLIAFRKAHPAL 658
Qy 578 NPKTFYALNSGTFYRSTDDGVTGPVAAAGLPSSGAVG-----VNFHVAVPKGE 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAVAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSVFTFLPAPSGTQWYRVTTDCDWDGASTFV-----APG 751
Qy 666 SSYPAVFVVGTTGGVGTGAYRSDDCGTTWVLINDQHQYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YQCGQSL 770
```









Search completed: May 11, 2004, 12:07:27  
Job time : 25.1745 secs

```

Query Match      3.6%; Score 145; DB 1; Length 747;
Best Local Similarity 20.3%; Pred. No. 0.36;
Matches 141; Conservative 69; Mismatches 234; Indels 252; Gaps 36;

Qy      179 DTTGYQSDIQGVVWVAFPKSSSLGQASKTITFVGADPNPVFWSRDGGAT----- 229
        ||| : : ||| : | : | : ||| : |
Db      55 DSTGKEAILSGVNWFGFN-----ASERVFHGL-----WS--GNITQTQQMAQR 96

Qy      230 -----WQA-----VPGATGTGP----- 242
        :| : |
Db      97 GINVRVPVSTOLLKWKAGTFLLKPNNVTYANPBLEGKNLSLQIPEYWLTLCORYIKVF 156

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 86.3468 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTPQYTSNVAIGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_25:\*\*

2: sp\_archaea:\*\*

3: sp\_bacteria:\*\*

4: sp\_fungi:\*\*

5: sp\_invertebrate:\*\*

6: sp\_mammal:\*\*

7: sp\_mhc:\*\*

8: sp\_organelle:\*\*

9: sp\_phage:\*\*

10: sp\_plant:\*\*

11: sp\_rodent:\*\*

12: sp\_virus:\*\*

13: sp\_vertebrate:\*\*

14: sp\_unclassified:\*\*

15: sp\_rvirus:\*\*

16: sp\_bacteriap:\*\*

17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2429.5	60.2	882	16 Q82M04	Q82M04 streptomyc
2	2371	58.7	890	16 Q86727	Q86727 streptomyc
3	2062.5	51.1	996	2 Q9AQH0	Q9AQH0 caldicellul
4	2009	49.8	839	16 Q97KK0	Q97KK0 clostridium
5	1765.5	43.7	857	3 Q8TFP1	Q8TFP1 aspergillus
6	1753.5	43.4	806	3 Q9P4T8	Q9P4T8 agaricus bi
7	1680	41.6	856	3 Q74170	Q74170 aspergillus
8	1579	39.1	838	3 Q7Z9M8	Q7Z9M8 trichoderma
9	1405	34.8	739	16 Q82K30	Q82K30 streptomyc
10	1162.5	28.8	751	16 Q8PLM5	Q8PLM5 xanthomonas
11	1160.5	28.8	707	16 Q9WVE1	Q9WVE1 thermotoga
12	1145	28.4	751	16 Q8P9U5	Q8P9U5 xanthomonas
13	1117	27.7	812	3 Q8J0D2	Q8J0D2 geotrichum
14	409	10.1	196	3 Q8J1H7	Q8J1H7 agaricus bi
15	368.5	9.1	1064	2 Q7X2U2	Q7X2U2 uncultured
16	238	5.9	2468	16 Q9I2M3	Q9I2M3 pseudomonas

17	214	5.3	1800	2	Q9L948	Q9L948 pseudomonas
18	214	5.3	8682	16	Q88RG2	Q88RG2 pseudomonas
19	205	5.1	2215	16	Q7WBN0	Q7WBN0 bordetella
20	202.5	5.0	2174	16	Q92U08	Q92U08 rhizobium m
21	198.5	4.9	3346	16	Q7WNS4	Q7WNS4 bordetella
22	197	4.9	2016	17	Q8TJ59	Q8TJ59 methanosarc
23	189	4.7	3972	16	P73139	P73139 synechocyst
24	186.5	4.6	4199	16	P74440	P74440 synechocyst
25	185.5	4.6	818	6	Q9N1P0	Q9N1P0 bos taurus
26	185.5	4.6	997	2	Q9Z4I1	Q9Z4I1 bacillus sp
27	184.5	4.6	998	17	Q9HL09	Q9HL09 thermoplasma
28	184	4.6	805	16	Q88GZ6	Q88GZ6 pseudomonas
29	183	4.5	2523	17	Q8TJ58	Q8TJ58 methanosarc
30	182.5	4.5	848	16	Q9RK65	Q9RK65 streptomyc
31	182	4.5	1070	16	Q8XW00	Q8XW00 ralstonia s
32	182	4.5	1904	9	Q38319	Q38319 lactococcus
33	179	4.4	2079	16	Q7UF08	Q7UF08 rhodopirell
34	178.5	4.4	3420	16	Q8FUS1	Q8FUS1 brucella su
35	177	4.4	666	11	Q8BRF6	Q8BRF6 mus musculu
36	177	4.4	928	11	Q80UN3	Q80UN3 mus musculu
37	177	4.4	1462	11	Q80WR0	Q80WR0 mus musculu
38	175.5	4.3	2523	16	O53393	O53393 mycobacteri
39	174.5	4.3	3930	16	Q98E20	Q98E20 rhizobium l
40	174	4.3	1271	17	Q8TI55	Q8TI55 methanosarc
41	172	4.3	993	16	Q8YYQ5	Q8YYQ5 anabaena sp
42	171.5	4.2	344	16	Q7W1M1	Q7W1M1 bordetella
43	171.5	4.2	7716	16	Q7UW28	Q7UW28 rhodopirell
44	171	4.2	908	16	Q8YR62	Q8YR62 anabaena sp
45	170.5	4.2	344	16	Q7WPL2	Q7WPL2 bordetella

## ALIGNMENTS

### RESULT 1

Q82M04	ID	Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;				
DC	01-JUN-2003 (TREMREL. 24, Created)				
DT	01-JUN-2003 (TREMREL. 24, Last sequence update)				
DT	01-OCT-2003 (TREMREL. 25, Last annotation update)				
DE	Putative endo-1,4-beta-glucanase.				
GN	CELA3 OR SAV1856.				
OS	Streptomyces avermitilis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=33903;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=21477403; PubMed=11572948;				
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,				
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,				
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;				
RT	"Genome sequence of an industrial microorganism Streptomyces				
RT	avermitilis: deducing the ability of producing secondary				
RT	metabolites.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=22608306; PubMed=12692562;				
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,				
RA	Sakaki Y., Hattori M., Omura S.;				
RT	"Complete genome sequence and comparative analysis of the industrial				
RT	microorganism Streptomyces avermitilis.";				
RL	Nat. Biotechnol. 21:526-531(2003).				
DR	EMBL; AP005028; BAC69567.1;				
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR001919; Bac celose-bind.				
DR	InterPro; IPR008965; Cellul bind.				
DR	InterPro; IPR002860; GH_BNR.				

```

DR Pfam; PF02012; ENR; 10.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD_II; 1.
KW Complete proteome.
SQ SEQUENCE 882 AA; 92152 MW; A3C8E753CE69F13D CRC64;

Query Match 60.2%; Score 2429.5; DB 16; Length 882;
Best Local Similarity 59.0%; Pred. No. 1.6e-126;
Matches 438; Conservative 111; Mismatches 174; Indels 19; Gaps 8;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRWIPILLDWGNNWGY 65
Db 39 YSKNARVDGGGFVPGIVFNESEKRLAYARTDIGGAYRWAESSKTWTLLDSVGVSDWGH 98
QY 66 NGVSVIAADPINTNKVMAAVGMYTNSWDPNDGAILRSSDQAGATWQITPLPFKLGNNMGR 125
Db 99 TGVVSLASDVDPNKVYAAVGTNTNSWDPNGAVLRSRSGRSGAQKTDLPFKLGNNMGR 158
QY 126 GNGERLAVDPNNNDNLIFGAPSGKGLWRSTDSGATWSQMTNPPDPVGTYIANPTDTTGYQS 185
Db 159 GNGERLAVDPNRSVLYLGAPSGKGLWRSTDSGASWSQVTFPNVGTIVQDADTDSGYAS 218
QY 186 DIQGVVWVAFDKSSSLQASQKTIFFGVADPNPNVFNWSDGATWQAVPGATGFIPIHK 245
Db 219 DNOGIVVWTFDESTGSGSSTRVTYGVADKDNSVYRSTDRAGATWSRLAQPTGHLAHK 278
QY 246 VFDPNVHLYIATSNITGCPYDSSGVKFSVTSCTWTRISGVPSTDTANDYFGYSGTLTI 305
Db 279 VLDRAAGCLYLAYSQKPGYDGGKQLWRYTKTGTNTNISPVAEADT---YFGSGTLTV 335
QY 306 DRQHPNTIMVATQISWPPDTTIFRSTDCGATWTRIDWTSYPNRSLRYVLDISAEPLWTF 365
Db 336 DRQHPGTWATAYSWPPDTQLFRSTDSGTTKAWDVTSPSRNRRFTMDVSSSPMLTW 395
QY 366 GVQPNPVPKLGWMDAMADPNRSDMLYGTGATLYATNDLTKWSDGGQIHTAPMVK 425
Db 396 GANPAPPQTPKLGWMTSLSLIDPPDSARMYGTGATVYGTNDLTKWSDGSGQTIKPMAR 455
QY 426 GLEETAVNDLISPPS-GAPLISALDGLGFTHADVTAVPSTFTSPVTTGTSVDYAEIN 484
Db 456 GLEETAVNDLASPPSGGQLFSLGDIGGFRHTDITVPSLMYTSFNITTSLSLYAETD 515
QY 485 PSIIIVRAGSFPSSQPNDRHVAFTSDGGKNNFQSGEPGVTGGTVAASADGRSFEVWAPG 544
Db 516 PGTVVRVGNLD--SGP---HVAFTSDGANFAGADPSVSGGTVAASDGRSFEVWSPA 570
QY 545 DPGQPVTVAVFGNSWAASQGVPAQAQTRSRVNPKTFFYALNSGTFFYRSTDCGVTFPQVA 604
Db 571 GTG--VOYTTGFTGTSWSASAGLPAGAIVESDRVDPKTFYFKSGRFYVSSDGGATFTASA 628
QY 605 A-GLPSSGAVGMFHAVPGKGDWLAASS-----GLVHSTNGGSSWSAITGVSSAVNVG 658
Db 629 ATGLPSSGDS--VRFKALPGTGDILWAGASDGAAYGLWHSTDGGAAFTKLAIVDQADTIG 686
QY 659 FGKSAPGSSYPVAVFVVGTVGAYRSDCCCTTWLINDQHQHYGNMGQALITGPHANLR 718
Db 687 FGKAATGASYQTLTYSAKIGVGRGIFRSTDKGASWTRVNDDAHQWGTGAALITGPRVYG 746
QY 719 RVYGTNGRGIVYDGGAPSG 740
Db 747 RVYVSTNGRGIVYDGTAGSSDG 768

RESULT 2
OB6727
ID OB6727
AC OB6727;
DT 01-NOV-1998 (TRENBERL 08, Created)
DT 01-NOV-1998 (TRENBERL 08, Last sequence update)
DT 01-OCT-2003 (TRENBERL 25, Last annotation update)
DE Putative secreted cellulase.
GN SC06545 OR SC5C7.30C.
OS Streptomyces coelicolor.

```

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2] SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3] SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN [4] Mol. Microbiol. 21:77-96(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RN Nature 417:141-147(2002).
DR EMBL; AL939128; CAA20642.1; --
DR PIR; T35237; T35237.
DR HSP; P07986; IEKQ.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR02860; CH_BNR.
DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD_II; 1.
KW Complete proteome.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 58.7%; Score 2371; DB 16; Length 890;
Best Local Similarity 57.7%; Pred. No. 2.8e-123;
Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRWIPILLDWGNNWGY 65
Db 41 YTKNARIDGGGFVPGIVFNESEKRLAYARTDIGGAYRWAESSKTWTLLDSVGVSDWGH 100
QY 66 NGVSVIAADPINTNKVMAAVGMYTNSWDPNDGAILRSSDQAGATWQITPLPFKLGNNMGR 125
Db 101 TGVVSLASDVDPNKVYAAVGTNTNSWDPNGAVLRSRSGRSGAQKTDLPFKLGNNMGR 160
QY 126 GNGERLAVDPNNNDNLIFGAPSGKGLWRSTDSGATWSQMTNPPDPVGTYIANPTDTTGYQS 185
Db 161 GNGERLAVDPNHDVLYLGAPSGHGLWRSTDSGATWSEVTAPENPGYADPDNDTSGYAS 220
QY 186 DIQGVVWVAFDKSS--SSLQASQKTIFFGVADPNPNVFNWSDGATWQAVPGATGFIPIHK 244
Db 221 DNOGIVVWTFDESTGSGGAGTATRTLYGVADKENVYRSTDRAGATWERLAGQPTGYLAHK 280
QY 245 GVFDPNVHLYIATSNITGCPYDSSGVKFSVTSCTWTRISVVPSTDTANDYFGYSGLT 304

```

```
Db 281 GVLDAENGYLAYSVDGPGYDGGKGLRYATATGTWTDISPAAEDT---YGFSGLT 337
Qy 305 IDRHQNTIMVATQISWMPDITIFRSIDGGATWTRIMDWTSYPNKSLRYVLDISAEPWLT 364
Db 338 VDRQPGTVMATAYSSWMPDITIFRSIDSGATWQMSYTSYDPRENYTMDVSSPWL 397
Qy 365 FGVQNPFPVPSKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKW-DSGGQIHIAPM 423
Db 398 WGANPAPPEQTPKLGWMTALEIDPFDSDRMYGTGATVYGTENLTNWDDEGGTFAVEPM 457
Qy 424 VKGLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIFTSVFTTGTSDVDAEL 483
Db 458 VRGLEETAVNDLAGPPSGAPLISALGDLGGTTHADVTAVPSTIFTSVFTTGTSDVDAEL 517
Qy 484 NPSIIVRAGSPDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGCTVAASADGSRFVWAP 543
Db 518 KPDVVVRAGNLD--SGP---HIAFTDNGANWFGTDPGSGVGGTVAAGADGSRFVWSP 572
Qy 544 GDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDDGGVTFQPV 603
Db 573 --EGAGVQYTTGFTSMQASTGLPAGALVESDRVNPATFYGKSGRFVYSTDGGATFTAS 630
Qy 604 AA-GLPSSGAVGVMFHVPKGEGLWLAASS-----GLYHSTNGSSWSAITGVSSAVNV 657
Db 631 AATGLPAGD--GVRFKALPGEGDVWLAGGAADGPYGLWHSTDDGGTFTRLPGVDAADTV 688
Qy 658 GFKSPAGSSYPVAVVVGTTGGVTVGVSDDCGTTWVLINDDOHQYGNWGAITGDHANL 717
Db 689 GFKAAPAGSYQTULTFSAIEGVNGIFRSTDAGATWTRVNDDAHQWGTGAATITGDPVY 748
Qy 718 RRVVIGTNGRGIYVGD I---GGAPSG 740
Db 749 GRVYVATNGRGIYVGD I---GGTDPG 775

RESULT 3
Q9AHO PRELIMINARY; PRT; 996 AA.
AC Q9AHO;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INEC.
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 51.1%; Score 2062.5; DB 2; Length 996;
Best Local Similarity 52.6%; Pred. No. 4.1e-106;
```

```
Matches 387; Conservative 96; Mismatches 238; Indels 15; Gaps 9;
Qy 4 QPYTWSNVAI-GGGFVDGIVFNEGAPGLIYRTRDIGMYRDAANGRWIPLLDWGWN 62
Db 35 QPYVKNVKIEGGGFTITGVFNPKKNLVYRTRDIGAYRSTGGNTWTQLMDWSFDD 94
Qy 63 WGYNGVSIADPINTKV-WAANGMYTNSWDPNDAIILRSSDQATWITPLPFLKGN 121
Db 95 WNLGVESIATDPVDPNVILACQGYTNSWTDMMGAILRSTDEGDTFFITPLPFLKGN 154
Qy 122 MPRGCMGERLAVDPNNDILYFGAPSGKGLWRSTDSGATWTSQMTNPPDVCTYIANPTDTT 181
Db 155 MPRXNIGERLAIIDPNNRILYLGREGKGLWSESDGVSWKKTSPFNPGTYIEDNCPN 214
Qy 182 GYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDSGATWQAVGAPTGI 241
Db 215 DYLNIHTGVWVVDPTSGRPGEGKIIYGVADKTTSIYYTKDGGCTWQALPQPTGLL 274
Qy 242 PHKGVDPVNVHLYIATNTGGPYDSSGDWVKFSVTSCTWTRISVPSTDTTANDYFGYS 301
Db 275 PQRKLSS-DGMLYITYTSNTQPYNGDYGEVWRNTKTGEWKNISPMQAQDT---YFGY 330
Qy 302 GLTIDRQHPNTIMVATQISWMPDITIFRSTDCGATWTRIMDWTSYPNKSLRYVLDISAEP 361
Db 331 GLAVDAQNPVKVWVAALSSWMPDITYIWRSIDGGETWKCIEWNGYPNRTLHYNNDISAAP 390
Qy 362 WLTFG-VQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Db 391 WLNFXETBTPPEVSPKLGWVGTLEIDPFNSDXMLYGTGAXLYGCCDLLTNWDKQNTI 450
Qy 421 APVKGLEETA VNDLISPPSGAPLISALGDLGGTFHADVTAVPSTIFTSPVFTTGTSDV 480
Db 451 KVKAIGIEETSQVALISPPVGPPLFSALGDIAGFRHEDLEKAPNWTYVQPNMGTITDIF 510
Qy 481 AELNPSIIVRAGSPDPSSQPNDRHVAFTDGGKWNFQSG-SEPGVITGCTVAASADGSRF 539
Db 511 AELNPSFVVRVGNVDKQWNPNTNRIGFSYDGGKSWFQGTGTEGSEGGTVAASADGSAV 570
Qy 540 VWAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDDGGVT 599
Db 571 VWAP--KGAKVCYSTDNGKNVCEANVPSEAIVYSDRVNPNKPYAFKNGKFIYISADKGT 628
Qy 600 F-QPVAAGLPSSGAVGVMFHVPKGEGLWLAASSGLYHSTNGSSWSAITGVSSAVNVG 658
Db 629 FIESPAAGLPISG---NFKTVPGIEGLIWLGVNGNMWHSITDGGYSFVKISGVDEDAASIG 684
Qy 659 FKSAPGSSYPVAVVVGTTGGVTVGVSDDCGTTWVLINDDOHQYGNWGAITGDHANL 718
Db 685 FGKPAEGETYPATYAKINGVIRGIFRSDDCDKTIRINDKXHQFGCANADITGDPVY 744
Qy 719 RVYICTNGRGIYVGD I 734
Db 745 RVFVATNGRGIYVGD I 760

RESULT 4
Q97KKO PRELIMINARY; PRT; 839 AA.
AC Q97KKO;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Probably secreted sialidase, several ASP-boxes and dockerin
DE domain.
DE CAC0919.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
```







01-JUN-2003 (Tremblrel. 24, Last sequence update)  
01-OCT-2003 (Tremblrel. 25, Last annotation update)  
Putative glycosyl hydrolase.  
SAV2574.  
Streptomyces avermitilis.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
NCBI\_TaxID=33903;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites";  
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005031; BAC70285.1; -;  
DR GO; GO:0016787; F:Hydrolase activity; IEA.  
DR InterPro; IPR002860; GH\_BNR.  
DR DR InterPro; IPR006311; Tat.  
DR Pfam; PF02012; BNR; 9.  
DR TIGRFAMs; TIGR01409; Tat\_signal\_seq; 1.  
DR Hydrolase; Complete proteome.  
KW SEQUENCE 739 AA; 78399 MW; 8E3B0F3D0F5B3D28 CRC64;  
Query Match 34.8%; Score 1405; DB 16; Length 739;  
Best Local Similarity 40.8%; Pred. No. 7.5e-70;  
Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;  
QY 6 YTSNVAIGGGGFDVGFVNEGAGFGLYVRDIDGMYRWDAANGRWIPLLDWGNNGY 65  
DB 42 YRWNAVIGGFTVGVLPFSPVGLAYARDIDGAYRWDRGARWPLIDHLGDDWNL 101  
QY 66 NGVSVIADPINTNKWAAVGMVYNSWDPNPDGAILRSSDQATWQITPLPKLGNMGR 125  
DB 102 LGVEAMAVDPHPRDLVLAAGTYAQSWAGN-GAVLRSEDGATWTRTDLTVKLGNEGR 160  
QY 126 GMGERLAVDPNNILYFGAPSGHGLWRSTDSGATWSQMTNPPDVGVYIANTDTTGVQS 185  
DB 161 GAGERLIVDRSDSTLWLT-RHDLGLKSTDRGATWAAATAPP-----AKANS 207  
QY 186 DIQGVVWAFDKSSSSLGQASKTIFVGVADPNP-----FVWSRDGGATWQVCAPTGF 240  
DB 208 SQGVWF-----IIVAGRTVYAGWGDGDSGTANLYRTAD-GTTWAGVGRSGT 257  
QY 241 ---IPHKGVDPVNHVLIYATNSGTGGVYDGSVDWPKFVSTGWTTRISPV-----PSTDT 293  
DB 258 SAKVPLRAAYDTHRELIVVYTDGAPGQSGSVHKLRTATGTWTEVTPVKPGGTTSDG 317  
QY 294 ANDYFGVSLTIDRQHNENTMVAQISWPPITIFRSTDDGATWRIWDTSYNRSURY 353  
DB 318 SADTFAYGVAVDARREGLTVNTRNWDGDTVFRSTDDGRTVTSKX-----AA 368  
QY 354 VLDISABEPLTFGVQPPVPSPKLGWDEAMADPFNSDRMLYGTGATLATNLTWTD 413  
DB 369 VFDVSETPFLDWGDD-----KPKFGWMIQALVDPYDSQHVYGTGATLTGTRDLKRW- 421  
QY 414 SGGQTHAPMVKGLEEAVNDLISPPSG-APLISALGDLGFTHADVTAVPST-IFTSPV 471  
DB 422 -----APRIRGLEEASVRLQISPPVGEAHLISGLGDIQVWYHERLTAASRGWATNPV 474

QY 472 FTTGTSVDYAEINPSIIVRAGSPDSSQPNDRHVAFTDGGKNW---FQSGPGCVTTGGT 529  
DB 475 FGSATGLAQAARPAFYVVRTGWDHNG-----AYSHDGGRTWAPFEAQPDIAKADPGP 528  
QY 530 VAASADGSRFWA-----PGDPQPVVYA-----VGFNGSWAASQGVPAQAIRSDRNPXT 581  
DB 529 IATSADGGTLLWSFVHWG-----TTAAHSTDNAGSMSEVSPKPGATPVADPADPTR 583  
QY 582 FYA--LSNGTFYRSTDDGGVTFQPVAAAGLPSSGAVGMFHAVPGKEGDLWLAAS-SGLYHS 638  
DB 584 FYAVDFDNGTLYASTDSGRSFTARAGLP-SGDSQFKLVAAPGRSGDLWSAKWNGLYRS 642  
QY 639 TNGSSWSAITGVSAVNVGFKSPSSYPAVFGVTGGTGGTAYRSDCGTTWVLND 698  
DB 643 TDGDTFARIDSCWASTYTLGFGKAADGADYPAIQVSGTETITATVYRSDAARTWRIND 702  
QY 699 DQHQYGNWQAITGDHANLRVYICNGRGIVYGD 733  
DB 703 DAHONGWIGEAUVGDPRIHGRVYLATNGRGIQYE 737  
RESULT 10  
QBPLM5 PRELIMINARY; PRT; 751 AA.  
ID Q8PLM5  
AC Q8PLM5;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Cellulase.  
GN CELA OR XAC1770.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=22024145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
Camarotte G., Cannavaro F., Cardozo J., Chambergo F., Ciapina L.P.,  
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,  
Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
Pereira H.A.F., Rossi A., Sena J.A.D., Silva C. de Souza R.F.,  
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
Nature 417:459-463(2002).  
DR EMBL; AE011809; AAM36634.1; -;  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 7.  
KW Complete proteome.  
SQ SEQUENCE 751 AA; 81259 MW; D4134D96285E599D CRC64;

Query Match 28.8%; Score 1162.5; DB 16; Length 751;  
Best Local Similarity 35.7%; Pred. No. 2.1e-56;  
Matches 268; Conservative 115; Mismatches 306; Indels 61; Gaps 20;  
QY 2 TTQPTWSNVAIGGGFVDGIVFNEGAGFGLYVRDIDGMYRWDAANGRWIPLLDWGN 61  
DB 43 TDDPYQMSVATGCGGFTVGVLPFAERDLAVARDVGGAYRWDAQAQVALTDWLGAD 102  
QY 62 NNGYGVVSIADPTNTNKWAAVGMVYNSWDPNPDGAILRSSDQATWQITPLPKLGN 121  
DB 103 DWNLTGIDAFVDPADPNALYLAAGTYMHERAGN-AAVLRSPDRGQFERADLPFKLGN 161



Qy	122	MPGEGMGERLAVDPNNNLLYFGAPSGKGLWRSTDGATWSQMTNFPDPVGTVIANPTDTT	181
Db	162	QLGRANGERLAVDPHDGRVLLGSRDA--GLWRSDDRGAHWARVEGFPADALAGATARNHV	220
Qy	182	GYSDIOGVVWVAFDKSSSSLGQASKTIFVGVDAPNNPVFWSRDCGATWQAVPGAPTGI	241
Db	221	GSEQAV--GIAFVFDAAAGHAGATPRIFVGVSTAQTSLYVSEDAGRWSAVAGQKGLR	279
Qy	242	PHKGVFPVNVHL-----YIATSNTPGPDGSGDVMKFSVTSGTWTRISFVPSFTDTA	294
Db	280	P-----SHMVGHNAQWYLSYCDRPGPDLMAGCALWNYDATQGRWEISPIQPAT-	330
Qy	295	NDYFGYGLTIDRHPNTIMVATQISWWPDYIIFRSTGGATWTRIDWTSYPNRSLRV	354
Db	331	GDGFGWGAVADPQHPQVLLASTFRRTPRDEFRSGDGRSWVPLLAQF-----	382
Qy	355	LDTSAEPWLTFGVOPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDS	414
Db	383	-DHSAAFW-TAHATPH-----WIG-ALAIIDPFDSNHATFVTGYGIWASRNLQTFS	430
Qy	415	GGQIHTAPMYKGLBETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTT	474
Db	431	QQPLQWFFQDRGLEETVPLDLLSPMAGAHLLSALGDIDGFRH-DALDTAQLQYLGRUTN	489
Qy	475	GTSVDVYAEIIMPSSIIVRAGSPDPSQPNDRHVAFTDGGKNWFQ--GSEPGVVTGTVAA	533
Db	490	GESIDGAGQAPQWVRSQTV--RDRRNNEIRALYSQDGGTHWAAFASEPPRGQAGTATA	548
Qy	534	ADGSRFWAPCDPCQPVVYAVG--FGNSWAAQGVPPANAIQRSDRVNPKTFYALS--	590
Db	549	ADASQVWVVPDQGG---VWRTGDFGKRWQVQGLPDTAVVAVDRDAQRWYAADRVSGL	605
Qy	591	YRSTDGQVTFQFVA--AGLPSGAVGVMFHAPVPGKEGDLMLAASSGLYHSTWCGSSWS	646
Db	606	YESDGAASFDTQOQVSPARDE-----RARPQLRPDPWRAGVYVILASPTLGVWRWQ	660
Qy	647	---AITGVSAVNVYFGKSAFGSSYPVAVFVGTIGVTCGYRSDCGTWTWVLINDHOHY	703
Db	661	QLRTLKSDEARSLGIGKALRAGAPPALLYLAGRVAGVDGIFRSDDGAWHRRINDDAHR	720
Qy	704	GNWQOAITGDHANLRVYVIGTNGRGIYGD	733
Db	721	PK--PYSVTGDPRIAGRVYFATGGRGIFYGD	749
RESULT 11			
Q9WYE1	ID	Q9WYE1	PRELIMINARY; PRT: 707 AA.
AC	Q9WYE1		
DT	01-NOV-1999	(TrEMBLrel. 12, Created)	
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	
DE	Endoglucanase, putative.		
GN	TM0305.		
OS	Thermotoga maritima.		
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
OX	NCBI_TaxID=2336;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MSB8 / DSM 3109;		
RX	MEDLINE=99287316; PubMed=10360571;		
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,		
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,		
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,		
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,		
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,		
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;		
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from		
RL	genome sequence of Thermotoga maritima."		
RL	Nature 399:323-329(1999).		
DR	EMBL; AE001712; AAC35393.1; -		
DR	PIR; F72393; F72393.		

OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OX	Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=340;	
[1]	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 33913 / NCPPB 528;
RC	MEDLINE=2202145; PubMed=12024127;
RX	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA	Canarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
RA	Cicarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA	Seubal J.C., Kitajima J.P.;
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing
RT	host specificities.";
RL	Nature 417:459-463(2002).
DR	EMBL; AE012276; AAM41043.1; --
DR	GO: 0004197; F: cysteine-type endopeptidase activity; IEA.
DR	GO: 0006457; P: protein folding; IEA.
DR	GO: 0006508; P: proteolysis and peptidolysis; IEA.
DR	InterPro; IPR001179; FKBP_PPIase.
DR	InterPro; IPR002860; GH_BNR.
DR	Pfam; PF02012; BNR; 7.
DR	PROSITE; PS00453; FKBP_PPIase 1; 1.
DR	PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
KW	Complete proteome.
SQ	SEQUENCE 751 AA; 81140 MW; 7378B6AA0B5698A4 CRC64;
Qy	Query Match 28.4%; Score 1145; DB 16; Length 751;
Db	Best Local Similarity 35.0%; Pred. No. 2e-55;
Qy	Matches 266; Conservative 121; Mismatches 295; Indels 78; Gaps 21
1	ATTQFTWSNVAIGGGFVDGIVFNEGAPGLYVRTDIGMYRWDAAANGRIPLLDWVGW 60
41	ATSGPYQWRSVAIGGGFVTGLFHPAERGLAYARTDVGAYRWDAQQWALTDLWGA 100
61	NNWYNGVSVTAADPINTNKWAAVGMVNTSNDNDGAILRSSOGATWQITPDPFKLGG 120
101	DDNLMGIDAFVDPADADALYLAAGTYMHERAGN-AAVLSFRNRRGTFERADLPFKLGG 159
121	NMPGRGGERLAVDPNDNITLYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYANPDTT 180
160	NQLGRANGERLAVDPHGRVLLLSGRDA-GLWRSDRGHAWKVASPPDAALAGATARNH 218
181	TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVGADPNPNPFWSRDGGATWQAVCAPTGF 240
219	VGREQAV-GIAFVVFDAASGNTGTPTRIYGVSTEQTSLYVSEDAGRSWAPVAGQPRGL 277
241	IPHKGVFDPNVNHLVYIATSN'TGGPYGSSGVDWKFSTVSGTWTTRISVPVSTDTANDYFGY 300
278	RPSHMAGSGDGH-WYLSYGDQPGDPLMAGGALWKFTPAQGRWRBISPIPQ-PASGDGFGW 335
301	SLGLTIDRQHPNTINVAQISWMPDTIIFRS'DGGATWTRIWDWTSYPNRSRYVLVDISAE 360
336	GAVAVDPQOPVOLLASTFRRTPRDELYRSVDGGKHAPL-----LADAVFDHSA 386
361	PWLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
387	PW-TAHATPH-----NWG-ALAI'DPFGNHALFVTGVGIWASRNL-----QDFA 428
421	AP-----MVKGLEATVNDLISPPSGAPLI'ISALGDLGQFTHADVTAVPSTIFTSPVF 472
429	APORPLQWFOORGLEETVPLDLLSPWAGAHLLSALGIDIGDFRDELDR-A-OLOVAGPRL 487



```

Qy 484 NPSIIVRAGSFPDSSQP-----NDRHVAFSTDDGKKNFQGS----- 519
Db 490 TPLV-----FSPTDPKTLVYGMQHVLRTNDGAVNTEISPDLTQKTTEPVKKNHADFDD 543
Qy 520 ---BPGGVTTG--GTVAASADGSRFVWAPGDPQPVVYVAVGFGNSW--AASQGVPAANAQI 572
Db 544 EEDKPQAPANGVIQTIAPSAKAGIIVWGTSTGLVQLTRDG-GKNWQNVTPPGLPERRAI 602
Qy 573 RSDRVNPK---TFYAL-----SNGTFYRSTDDGVTQFQVAAAGLPSSGAVGVMPHFAVPGK 623
Db 603 ILIEASPRNADTAIVIGVAIPDSHPYIYRTDGGKSWQKIVTGLPKGIARVV-REDPTR 661
Qy 624 EDDLMLAASSGLYHSTNGGSSWSAI 648
Db 662 EGLVYAGTETGAHVSYDGDHWQTL 686

```

Search completed: May 11, 2004, 12:09:26  
Job time : 90.3468 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 36.3141 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWMSNAIGGGGFVD.....YIGTNGRGIVYIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.\*
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219.5	5.4	2736	4	US-09-252-991A-30227
2	181.5	4.5	3892	4	US-09-328-352-5503
3	162.5	4.0	772	2	US-08-410-784A-5
4	158.5	3.9	774	3	US-09-346-237-8
5	157.5	3.9	688	4	US-09-489-039A-9813
6	154	3.8	1548	4	US-09-232-991A-22301
7	152.5	3.8	1751	3	US-09-136-574A-44
8	152	3.8	3472	4	US-09-408-020-4
9	151.5	3.8	776	3	US-09-346-237-4
10	148.5	3.7	750	6	5457037-3
11	148.5	3.7	751	6	5457037-5
12	148.5	3.7	776	3	US-09-346-237-7
13	147	3.6	3623	4	US-09-341-461-2
14	146	3.6	1749	4	US-09-640-419C-28
15	143	3.5	894	3	US-08-362-525-22
16	143	3.5	894	3	US-08-971-692-15
17	142	3.5	1216	4	US-09-134-000C-5130
18	141	3.5	720	3	US-09-236-284-25
19	141	3.5	754	3	US-09-236-284-4
20	140.5	3.5	387	4	US-09-252-991A-21962
21	140.5	3.5	1000	3	US-09-352-159-25
22	140.5	3.5	1000	3	US-09-352-168-25
23	139	3.4	826	4	US-09-328-352-7515
24	137.5	3.4	1298	4	US-09-252-991A-30579
25	137	3.4	619	4	US-09-252-991A-26352
26	136	3.4	2628	2	US-08-570-311-14
27	135.5	3.4	1205	3	US-09-352-159-29

```

28 135.5 3.4 1205 3 US-09-352-168-29
29 135.5 3.3 1295 4 US-09-328-352-6045
30 132.5 3.3 626 6 5268290-2
31 132.5 3.3 866 1 US-08-386-727-8
32 132.5 3.3 866 2 US-08-600-452A-8
33 132.5 3.3 1395 3 US-09-540-245A-15
34 131.5 3.3 1026 2 US-08-542-003-6
35 131.5 3.3 1026 2 US-08-322-760A-6
36 131.5 3.3 1026 4 US-09-236-949-6
37 131 3.2 1638 4 US-09-071-035-258
38 131 3.2 1638 4 US-09-071-035-262
39 131 3.2 1638 4 US-09-071-035-266
40 131 3.2 1747 4 US-09-134-000C-5999
41 130.5 3.2 862 1 US-08-325-267A-4
42 130.5 3.2 1277 4 US-09-397-885-3
43 130.5 3.2 1277 4 US-09-969-362-3
44 130.5 3.2 1326 4 US-09-147-236-5
45 130.5 3.2 1326 4 US-09-522-474-5

```

#### ALIGNMENTS

```

RESULT 1
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

```

```

Query Match Similarity 5.4%; Score 219.5; DB 4; Length 2736;
Best Local Similarity 21.2%; Pred. No. 4.8e-07;
Matches 190; Conservative 94; Mismatches 324; Indels 289; Gaps 48;

QY 59 GWNNGY-----NGV--SIAADPI-NTNKVA-----AVGMYTNSWDPNDGAILR-SS 103
Db 1305 GGNWTFPTPATPLANGTVVNAVAQDPAGNTPGQSTTVDAVAPNTPVVPNSGNLLNGTA 1364

QY 104 DQATWQITPLPKLGGNPMGRGMRGLAVDPNNNDILYFGAPSGKLMRSTDSGATWSQ 163
Db 1365 EFGSTVTLT-----DGN--GNPIGTTA-----DGSNGWSFTPG-----SQ 1398

QY 164 MTNFPDVGTIYI-ANTDTTGYQSDTQGVVWVAFDKSSSLGQ---ASKTIFVGVDNPNP 219
Db 1399 LPN-----GTVVNVTASDAAGNTS---APATTVTSSLPISIPQVDPNSGVISGTADAGNT 1451

QY 220 VF-----WSRDGGATWQAVPGAP-----TGFI 242
Db 1452 IITDGNNGPIQVQVADGSGNWSFTPIGIPDGTVVNVVARSNSVDSAPAVITVDGVAP 1511

QY 243 HKGVDPVNVHLYIATSNITGGPY---DGSSGDVWFKEFSVT-SGTWTRISVPV-----289
Db 1512 AAPVIDPSNGTISGTAEAGATVILTDGGNPIGQATADSGNWTFTPGTPLANGTVINA 1571

QY 290 -STDANDYFGVSGLTIDRQHENTIMVATQISWPPDTIIFRSTGDTGATWTRIMDWTSPN 348
Db 1572 VAQDPAGNTSGPASTVTDIAIP-----PAPVINPNSGVISGTAEAGAT-VILTDGNPNP- 1625

```



```

; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,784A
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heine, Ph.D., Holliday C
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: ISU-002XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-2290
; TELEFAX: 617-451-0313
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-410-784A-5

Query Match 4.0%; Score 162.5; DB 2; Length 772;
Best Local Similarity 19.8%; Pred. No. 0.0012;
Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;

Qy 51 WIPLLDVGNNGYN---GVVS---IAADPINTKVAAGVMTNWDNDGAILRSSD 104
Db 108 WGP--NWPASNMGKSGAGCVSDVDANGDRFNPKLLDPYAQEBSDP----- 155
Qy 105 QGATQWITPLPFKLGGMPCGMGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQM 164
Db 156 -----LNPSNQN---GNVFASAHYRTDSGI----- 178
Qy 165 TNPDPVGTIYANPTDTTG-----YQSDIQGVVWVAFDKSSSLG----- 203
Db 179 --YAPKGVVLVPSTQSGTGTETKTRAQKDDVIYEVHVG-----FTEQDTSIPAQYRGTYG 231
Qy 204 ---QASKTIFVGVA-----DPNNPV-----FMSRDGGATWQA 232
Db 232 AGLKASYLASLGVTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNK 291
Qy 233 VPGAPTG-FIPHKGVFPVNHVLYIATNTGPGPYDGGSGDVWKFVSTGWTIRISPVST 291
Db 292 AAGGPTAEFOAMVQAFNAGIKVYMDV-----VYNHTABGGTWTSSDPTTAT 338
Qy 292 -----DTANDYF-GYSGLTIDROHPNTI---MVATOISWNPDTIIFRSTD 332
Db 339 IYWRGLDNATYELTSGNQYFYDNTGIGANFTNTYVAQNLIIVDSVAYWANTW---GVD 395
Qy 333 GGATWTRIMDWTSPNRSRLRYLDISAEPLMTFGVQPNPPVPSPKLGMWDEAMAIDPFNS 392
Db 396 GFR-----FDLASVLGNSCLNAVHASA-----FNCPNGYFNDAADSNVAI----- 436
Qy 393 DRMLY-----GTGATLY-----ATNDLTWK--DSGGQIHIAPMVKGLESTANDL 435
Db 437 NRILREFTVRPAAGGTVMICLRNLGPSAATR--TSWYDSRRVRVWVSVPRQLRAQNEL 494
Qy 436 ISPPSAPLISALGDLGFFTHADVTAPVSTIFTSPTFTGTSVDYAEINPSIIVR----- 490
Db 495 -----GSMTIYVTOANDFSGS-----SNLFQSGGRSPNNSINFIDVHDGNTLKDVSC 543
Qy 491 --AGSPDPSPQPNDRHVAFTSDGKNWFGQSEPGVTTGGTV---AASADGSRF-VWAPG 544
Db 544 NGANNQASVGPSDG-----GTSTNYSWDOQMSAG---TGAVDQRRRAARTGMAFEMLSAG 596

; 545 DP---GQPVVYAVGFGN-----SWAASQG---VPAQAQIRSDR---VNPX 580
; 597 TPLMQGDEYLRITLQCNNNAYNLDSSANWLITYSWTTDQSNFYTFQAQLIRSAHPILRPS 656
; 581 TFYALSNGTFRSTDDGGVTFQFVAAAGLPSGGAVGVMFHAPGKGBDLWLAASSGLVHSTN 640
; 657 SWYSGSLQWTY-----QF-----SGAV-----AUSNYWNNTS 683
; 641 GGSWSAITGVSSAVVWVGFKSAPGSSYPAFVFW--GTIGGVTGAYRSDDCGTWVLIND 698
; 684 NYATAYANGPSL-----GDSNDSIYVAYNGWSSSVTFTLPAPPSGTQWYRVTD 732
; 699 DQHQYGNWQAITGDHANLRVYICNGRGIVYGDIG 735
; 733 T---CDWNDGASTFVAPGSETLIG--GAGTTYGQCG 763

RESULT 4
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Favobacterium odoratum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(774)
; OTHER INFORMATION: Isoamylase
; US-09-346-237-8

Query Match 3.9%; Score 158.5; DB 3; Length 774;
Best Local Similarity 18.8%; Pred. No. 0.0025;
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;

Qy 76 INTNKVMAA-----VGMVYNSWDPNDGAILRSSDQATWQITPL 114
Db 35 INPNKLGAAAYDATKANVTFKVYSSKATRIELIYLYSTATGSAEKAKYVMTNSGGIWSVT-I 93
Qy 115 PFKLGNNPGRGMGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYI 174
Db 94 P---TSTLSGGQLG-----GTLYYGYRAWGNW---PYNASWTKGSSL----- 130
Qy 175 ANPTDTTGVQSDIQG-----VWVAFDKSSSLGQAKIIFVGVADNPP 219
Db 131 -----GFISDVAAGNRFNPNKLLSDPYALELSDPTTATW-----TNGS 170
Qy 220 VFWSRDGGATQAV---PGAPTGF1-----PHKGVFDPV---NHLVYIATNTG 262
Db 171 IYAS---GATVNIIDSGSAPKGIVLAGDTQATGTPTKALKDDVYEAHVRGLTMDTS 227
Qy 263 -----GPYDSSGGDVWKFVSTSGTWTRISPV-----PSTDTANDYFGYSGLTI 305
Db 228 ITAAARGYTKAGLKAALAAALGVTAIEFLPVQETQNDTNDNDPSSTSGDNWYGYMTLNY 287
Qy 306 ---DRQ-----HPNTIMVATOISWNPDTIIFRSTDGATWTRI 341
Db 288 FAPDRRYAYDKTPGGPTREFKEMVKAFHNDNGIKVLVD-----VYVNHTEGEGANSPT- 339
Qy 342 DWTSPNRSLR-----YVLDISAB-PWLTFCVQPNPPVPS-----PKLGMWDEAM 385
```



Db 340 DKTINTTSFGLONPTTYSITADPQNSMDNTVCGNTNTNTTIAQNLIVDSLAYWRDKL 399  
Qy 386 AIDPFNSDRMLYGCATLYATNDLTWKDSCGQIHIAPVKGLEETAVNDLIS--PP----- 439  
Db 400 GVDGYRDLASVIGNSCOHGFNFKMDAG-----NALNRIVAELPPRAT 445  
Qy 440 --SCAPLIS--ALG-----DIGGF-----THADVTAVPSTIFTSPTFTTGSVDYA 481  
Db 446 GSGVDLTAEPWAIGNSYOVGGFPSCWAENWAGYRDVROAKNGKLSVAITTG----- 499  
Qy 482 ELNPSIIIVR-AGSDPPSSQPNDRHVFSTDCGKWFQG---SEPGVTTTGTGTVAAASDGS 537  
Db 500 ----QMATRFAGSSD-----LYGDDGRKPMHVSNFITAHDGFTLKDLYSNCNKN 545  
Qy 538 RFVW--APGDPGQPVVYAVGFCNSW-----AASQ-----GVP-----ANAQ 571  
Db 546 NQWPFYPSDGE-----DNNNSWDQGGIAADQKARNGMALMLLSAGVPMIVGGDEA 599  
Qy 572 IRS-----DRVNPKTFY-----ALSNCTFYRSTD-G 596  
Db 600 LRSNMCNNPNYLDSSANLWNSRTTDQNNFQSFASKAMIAFRKAHPALRPANFYSSVDNN 659  
Qy 597 GVTQPVAAAGLPSSCAVCMF-----HAVPK-EGDLWLAASSGLYHSTNGGSSWSAIT 649  
Db 660 GNVNEQLRWFKPDGADVADATYFNDANNHAIAMRIDGSEFGDTASAIYVAHN---AWSA-- 714  
Qy 650 GVSSAVNVGFKSAPGSSYPAVFVVGTTGGVTVAYRSDCCTTWLINDDQHQVGNWQ 708  
Db 715 -----QVNFILPWFAGKSWYRVTIDCGWAGASQVQAPGE-ALVGENTAYGLCR 766

RESULT 5  
US-09-489-039A-9813  
; Sequence 9813, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489.039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9813  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9813

Query Match 3.9%; Score 157.5; DB 4; Length 688;  
Best Local Similarity 22.6%; Pred. No. 0.0025;  
Matches 111; Conservative 53; Mismatches 184; Indels 143; Gaps 24;  
Qy 60 WNWNG--YNGVVSIAADPINTNKVAAVGMVYTSNDPNDGAILRSSDQATWQITLPLFK 117  
Db 171 WAHGNTHGDRFAALDQNKQ-----VNQLQVAVHATGDIPOSNGSGAEDQNTPLQ-- 224  
Qy 118 LGGNMPGRGMRGLAVDPNNDNIIYFGAPSGKGLWR--STDGATW-----SQM 164  
Db 225 -----IGDTLYVCTPYSKVLADVSGEKWRVYDSKSSPNWQRCRGLGYVADSOA 275  
Qy 165 TNFPDVG-----YIANPDT-----TGYSOI---QGVVWVAFD 196  
Db 276 QTAPASGTQPAACSRRLFLPTIDARLIAIDADTKLCENFGDGGIVDLISVGMGEVKAGY 335  
Qy 197 KSSSSLQASKTIYFG--VAD---PNNPVFWSRGGATWQAVGAPTGFIPHKG-----VF 247  
Db 336 QQTSTPLVAGNVVVGKADVNTGEP-----PGVVRADFVHTGKLAWAW 381  
Qy 248 DPNVHLYIATSNTPGPGYDSSGDWKFVSFTSGT---WTRIS-----PVPSTDTAND 296

Db 382 DPGNPAL-----TGVPPEQ-----TYTRGTPNWSAMSADAKNLIYLTGNATPD 428  
Qy 297 YFGVSGLTIDRQHPNTIMV--AT--QISWMPDTIIFRSTDDGATWTRIMDWTGYPNRSLR 352  
Db 429 FGGERTALDDKYSSSIIVAVDATTGQVRWHFQT-----THDLWDF-DLPQPLL 477  
Qy 353 Y-----VLDISAEPWLTFTGQVPPNPVSPKLGWMDMAIDPPN-----SDRM 395  
Db 478 YDLPDGKGTTEVLVQTSKQGMIFMLNRETGEPVAKV---ERPVPAGNVKGERYSPTQ 533  
Qy 396 LYGTGATLYATNDLTKWDSGGQIHIAPVKGLE--ETAVNDLISPPSGAPLISALGDLGG 453  
Db 534 PYSVGMPIGNQTLTESDMWGATPIDLLCRIQFKEMRHQGVFTPPGEDRSLOFPGLSGG 593  
Qy 454 FTHADVTAVPS 464  
Db 594 MNWGSVSLDPN 604

RESULT 6  
US-09-252-991A-22301  
; Sequence 22301, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22301  
; LENGTH: 1548  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22301

Query Match 3.8%; Score 154; DB 4; Length 1548;  
Best Local Similarity 20.7%; Pred. No. 0.014;  
Matches 188; Conservative 87; Mismatches 270; Indels 364; Gaps 53;  
Qy 73 ADPINTNKVAAVGMV---TNSWD-----PNDGAILR----- 101  
Db 319 SDGVRQGRNWAELAEMLPAGSQSWSMRLVAGADLAAADNRLVRPDSASLADTHYQAKI 378  
Qy 102 -----SSDQATWQITLPLFKLGGNMPGRGMRGLAVDPNNDNIIYFGAPSGKGLWR 153  
Db 379 EQSSGLVFTDQATWGITP-----GTPVDESNEWI-----CGLGPYC 416  
Qy 154 STDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKS---SSLSGQASKTIFV 211  
Db 417 AEPPTWAP-----GNYLCPAGTAIGEGLW---MCSVDPSLCIENLGT----- 460  
Qy 212 GVADPNPNPVF-----WSDRG--GATWQAVPGAPTGFIPHKGVFDP 249  
Db 461 -VVTQNLFLSVLRTGTGDLDLASAGNLQWSPYGVYTAGTQAADVA--TCFNQPRGLFN- 517  
Qy 250 VNHVL-----YIATS-----NTGGPYD---GSS--GDVWKFVSFTGWTTRISVP 289  
Db 518 -GSVLGAGADYEVLSTSQYQAWPEHGNLNDIAVGDDVVGQWAEKLTSS--DPIRLPL 574  
Qy 290 STDANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIF-RSTDGATWTRIMDWT--- 344  
Db 575 PSAAVGNLWLRQG--SADREGVT-----ANWVFGSVRGAEGDAPY--LVGFTGFGT 624  
Qy 345 -----SYPN-----RSRYVLDISAEPLT----- 364  
Db 625 LGGNLSMRTGDNAGNIAPRGDSIPSSGNLNPSSQGLVLA VAGTGRLTSDGALQLGCGG 684

Qy 365 -----FGVQNPVPSPKLGWDEAMADPFNSDRM--LYGTG-----AT 402  
 Db 685 DLNVRIGGEVNP-----SREARATQYSSGDFGLYSGGTIHDLOQALINLRGSAS 735  
 Qy 403 LYATNDLTKWDSGQIHIAP-----MVKGLEETA VNDLISP-----PSGAPLIS-----446  
 Db 736 LY-----SGALGIDPRYDTLLRDPAEVRSRDAFPTLASSTGGLTLVAGDTGMR 786  
 Qy 447 -ALGD--LGGFTRADVAVPSTI-FTSP-----VFTTGTSDVDAELNPSII 488  
 Db 787 ETRGDLVLGGVDPGRGVGVNTVGTFTAPDGSVYGGGIGWFSLTWHTSID-----LP 839  
 Qy 489 VRAGSFPDSSQ-----PNDRHVAFTSD-----GGKNWFGGSE 520  
 Db 840 AAGNLTPSTQLEATNAIPWAGNLSPSDGRFIYPSIVRAAAPEGSYILGPSSGYMGV 899  
 Qy 521 PGVYTT---GGTVNASADGRFVWAPDGPQPVVYAVFGNSWASQGVPAQAQI-----572  
 Db 900 SLNVSTTPYSLLAPSUNGELLEIA-GDS-----IYAGGY--SVQSGADPANLPSIWT 952  
 Qy 573 ---RSDR--VNPKTFYALSNGTFYRSRDGGVTFQPVAAAGLPSSGAVGMFHAVPGKEGDL 627  
 Db 953 FAGYSDAALNP-----IAGNG-----SPDGN---PAVIGGLP-----LFFGPDASA 994  
 Qy 628 WLAASSGLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGVGTAYRSD 687  
 Db 995 ARDLQARFYALTGCD-----IVGLNSGAQIRFGEQA-----GNRAGOTWYE 1035  
 Qy 688 DCGTTWVLINDHQYQN-WQAI-----TGD---HANLRVYIGTNGRGIYV 731  
 Db 1036 GAGPVMWRAGRDIVASGTPLQRIAPSQISTDASFTGNLFPVHDDPNLDSLVAQGRDILY 1095  
 Qy 732 GDIGGAPSG 740  
 Db 1096 GNFNVAGPG 1104

RESULT 7  
 US-09-136-574A-44  
 ; Sequence 44, Application US/09136574A  
 ; Patent No. 6294366  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farrington, Graham K.  
 ; Anderson, Paige  
 ; Gibbs, Moreland  
 ; Bergquist, Peter  
 ; Daniels, Roy  
 ; Morgan, Hugh W.  
 ; Williams, Diane P.  
 ; TITLE OF INVENTION: Compositions and Methods for  
 ; Treating Cellulose Containing Fabrics Using Truncated  
 ; Cellulase Enzyme Compositions  
 ; NUMBER OF SEQUENCES: 49  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Howson and Howson  
 ; STREET: Spring House Corporate Center, P.O. Box 457  
 ; CITY: Spring House  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19477  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM compatible  
 ; OPERATING SYSTEM: DOS  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/136.574A  
 ; FILING DATE: 19-Aug-1998  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/932,571  
 ; FILING DATE: September 19, 1997  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.  
 REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: 1997US001/CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9200  
 TELEFAX: 215-540-5818  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 44:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1751 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:  
 US-09-136-574A-44

Query Match 3.8%; Score 152.5; DB 3; Length 1751;  
 Best Local Similarity 18.7%; Pred. No. 0.021;  
 Matches 184; Conservative 101; Mismatches 349; Indels 351; Gaps 49;  
 Qy 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTD-----IGMYRW-----D 45  
 Db 233 YTAATGFTSGFDIDLGW---AAWLYIATNDSSYLTKAEELMSEYANGTNTWTQCWDD 289  
 Qy 46 AANGRIWIFLDWGMNMGVNGVVSIAADPINTNKVAAVGM-YTNSWDPNDGAILRSSD 104  
 Db 290 VRYGTLIMAKITGKEL--YKGAVERNLHDHTDRITYTPKGMAYLTGM---GSLRYATT 343  
 Qy 105 QG-----ATWQ-----ITPLPKLGNMPPGR-----GMGERLAVDPNNNDNI 140  
 Db 344 AAFACVYADVADSGCDNSNKKTKYLNPAKSIDYALGST--GRSFVVGFGTNPQPHHRN- 400  
 Qy 141 LYFGAPSGKGLWRSTDSCATWSQMTNFPD-----VGTIANPTDTTGYSQDIQGVVW-- 192  
 Db 401 -----AHSSWANSMKIPEYHRHILYALVGGPGSDSDNDIDITDYVQNE 444  
 Qy 193 VAFDKSSSLGQASKTIFGVADPNPNVFWSDGGATQAVPGAPTGFPHKGVPDPVNH 252  
 Db 445 VADYNAGIVGALAK-----MYQLYGEPID--DFKAIEPTND 481  
 Qy 253 VLYIAT--SNTGGPYDSSGDVWKFVSVTSGTWTRISPVPSTDTAN--DYF-----GY 300  
 Db 482 EIFVESKEFGNSQGP---NVTIVISYIYNRTGM---PPRVTDKLSPKYFIDLTELIAQY 534  
 Qy 301 SGLTIDROHPNTIMVATQISWHPDIIIFRSTDGGATWTRIMDWTSYPNLSRVVL--DISA 359  
 Db 535 S-----PDVVKVDT-----YYIEGKISGPYVWD---KNRNIYVVLVDFSG 572  
 Qy 360 EPWLTFFGVQNPVPSPKLGW---MDEAMADPFNSDRMLYGTGATLYATNDLTKWDSG 415  
 Db 573 TK-----IYPGEVEHKKQAQPKISVPQGYPDPTN--DPSYKGLTSQLEKNKYIAAYDNN 626  
 Qy 416 GQIHIAPMVKGLEETA VNDLISP-----PSGAP-----443  
 Db 627 N-----LVWGLEPGAATSTPAPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 680  
 Qy 444 -----LISALGDL-----GGFTHADVAV-----P 463  
 Db 681 VLYKNNETSASTGSIRPWFKIVNGGSSVDLSRVKIRYWTYVDGKPSQAVCDWAQIGAS 740  
 Qy 464 STIFTSVPVTTGTS--VDYAEALNPSIIVRAGSPDPSSQPNDRHVAFTSDGKN-----514  
 Db 741 NVTNFKVLSGSGVSGADY--YLEVGSSGAGLOQPKDGTGDIQVRFNKNWNSYNTQADDDS 799  
 Qy 515 WFO-----GSEPGVTTGGTVAAASADGSRFVWAPGDP-----546  
 Db 800 WLOSMNTYGENAKVTLYVDGVLVWQEPGAPATPATSTATPTPTPTPTPTPTPTPTPTPTPT 859  
 Qy 547 -----GQPVVYAVFGNSWASQGVPA-----NAQIRSDRVNPKTFYALSN--GT---589  
 Db 860 VSATPTAPTASPVG--GSYWTSPESYGALKWYANGNLSPTNVLNPK--IKIENVGTGA 916

QY 590 -----FYRSDGGVT-----PQP-----VAAGLPSSGA-----VG 614  
Db 917 VLSRVKRVYITIDGEATQSVASSINPAYIDVKFKLGANAGGADYYVEIGKSGAG 976  
QY 615 VMFHAVPKGEGDLMAAASGLYHSTNGGSSWSA-----ITGVSSAVNVYFGKSAFSG 666  
Db 977 VLAAGQSTKEIRLSIQKSGSYNQSDYSVRSANSYIENEKVYTGVIDDVLVWGREPGRNA 1036  
QY 667 SYPAVFVVTGGVTGA-----YRSDDCGTT-----WVLNDHQHVGWNGQAIT 711  
Db 1037 QIKVWYANGNLGSMTNVLNPKIKIENVGTAVLSRVKRVYWTIDGATQSVSTSSIN 1096  
QY 712 GDHANLRVYITNGRGI-VYGDIG 735  
Db 1097 PAYIDRVFKLGANAGGADYYVEIG 1121

## RESULT 8

US-09-408-020-4  
; Sequence 4, Application US/09408020  
; Patent No. 6632937  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOOP.002A  
; CURRENT APPLICATION NUMBER: US/09/408,020  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 3472  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-09-408-020-4

Query Match 3.8%; Score 152; DB 4; Length 3472;

Best Local Similarity 21.0%; Pred. No. 0.058;

Matches 172; Conservative 100; Mismatches 244; Indels 304; Gaps 50;

QY 96 DGAILRSSDQATQWITP-----LPPKLGNNMPGRGMGERLAVDPNNDNILYFGAPSGRGL 151  
Db 986 DGTVVFLSD-GPTGRVYPSYSLNIPEDISSAAPG-----GFVIVPGVSDIAFSA-DGRNM 1038  
QY 152 WRSTDGATWSQMTNFP-DVG-----TYIANP--TDTTGYQSDIQGVVWVAF 195  
Db 1039 LVADETGGIHRYLARSPEYIGTDFIKSSLGBFVETFSAPRVQDLAGIAFSDGIMILAA 1098  
QY 196 DKSSS-----SLGQASKTIFVGVADPNPNPVFWS-----RDGGATWQAVPG- 235  
Db 1099 GGSGSVIRYSLPSYAVSGAKYEETAMIG-GSPGLEFSSDGLRMFVDPDAGSETAAVYGL 1157  
QY 236 -APTGFIPHGKGVDFV-----NHVLIATSNITGSP----- 264  
Db 1158 AAPYGI-----GEAEPLPLFLGVGAEEATLSPDGRHILV-----PGRPLSQYSLFSTNL 1208  
QY 265 -----YDGSDD--VWKFVSVTSGTWTRISPVSTDTAN-----DYFGVSG----- 302  
Db 1209 ELCAEPRGIDGSCCEDGIYAFE-SPGRGEGVSLAASITAADGPGIGELHGFAGPMPAPV 1267  
QY 303 ---LTIDRQH-----PNTIMVAT---QISWMPDT----- 1157  
Db 1268 MEQVTLDSREGTLRVLRDTRVDVTRPYKMWVEDSDGSGQTTLANSTLLNNAENSNILLFR 1327  
QY 330 STDGGA-----TWTRIW---DWTGYPNR-----SLRYVLDISAEPMLTFGVQ 368  
Db 1328 LDDAAAGKISGYTSFVFTNSSPFLGTGATRPHTLFGDVRADIIYDASG----- 1379  
QY 369 PNPVPVSPGK-LGWMDAEMADIPFNSDRMLYGTGA-----TLYATNDLTKWDSGGQIHIA 421  
Db 1379 PNPVPVSPGK-LGWMDAEMADIPFNSDRMLYGTGA-----TLYATNDLTKWDSGGQIHIA 421

Db 1380 -----VPSPSGIEFSDGMRM-----FVTGIGTGPCINIFTLSAPPDITILPKHSGSTNIG 1428  
QY 422 PMVKGLEETAANDLISPPSGAPL-----ISALGD-----LGGFTH-----ADVT-A 461  
Db 1429 ---GL---SVSDLAFANNNSLTVLVDVGLRVYALGDDYNNVVTGTTQKFRITLDTTQG 1481  
QY 462 VPSITFTSP-----VFTTGSVDYA-----ELNPSIIVRAGSFDPS-----QP 500  
Db 1482 IPNSIYTSPDGLSQFVAYDDRIDLVLGSPNDISSTTEIIPYSLPRP---DPPTGMDFTP 1538  
QY 501 NDRHVAFSTDGKNNFQSGEGVTTGG-----TVAASADGSRFV-----WAPGDPG-- 547  
Db 1539 DGRMFLSTENGIDQYLLSEPFVAVTTSVFLRTIPIDGGAGIRFVONGRGLFVPGADGII 1598  
QY 548 --QPVVYAVFGNSW--AAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTGDTGVTTPV 603  
Db 1599 QRHELIYPYGASTSLLETVRDGV-----TDGGPGENPA 1631  
QY 604 AAGLPSSGAVGMFHAVPKGEGDLMAASGLYHSTNGGSSWSAITGVSSAVNVGFKSA 663  
Db 1632 AGEIRLAGT-----FNASDN-----VQSPSGIEFSGDGTGMF--VTGFGAAGVNEFSLSA 1679  
QY 664 P-GSSYPAPVAVVVTGIG--VTGAYRSDDCGTTWVLNDD 699  
Db 1680 PFTLPLVHVELHDIGGQPAVDLAFABD--GRTLILLAAAD 1717

## RESULT 9

US-09-346-237-4  
; Sequence 4, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:  
; APPLICANT: Bisgaard-Prantzen, Henrik  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629.200-US  
; CURRENT APPLICATION NUMBER: US/09/346,237A  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER APPLICATION NUMBER: PA 1998 00868  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: 60/094,353  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 776  
; TYPE: PRT  
; ORGANISM: Pseudomonas amyloclavata  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(776)  
; OTHER INFORMATION: Isoamylase  
US-09-346-237-4

Query Match

Best Local Similarity 3.8%; Score 151.5; DB 3; Length 776;

Matches 171; Conservative 89; Mismatches 274; Indels 291; Gaps 47;

QY 27 GAPGLYVRTDYGMYRWDANGRIPLLDVGVNWNWNGVVSIAADPINTKVMVAVG 86  
Db 96 GITGAVY-----YGRWNGPNWYASNWKGSQAGFVSDVDANGDRFNPKLLDPY 147  
QY 87 MYTNSWDNDGAILRSSDQATQWITLPLFKLGNMPPGRGMGERLAVDPNNDNILYFGAP 146  
Db 148 AOEVSQDP-----LWPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182  
QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFKSS--SS 201  
Db 183 KGVVLVPSTQSTGTXFTRAKQDDVIEVHVGRFTQEDTSPAQYRTGYAGLKASVLAS 242  
QY 202 LQQASKTIFVGVAD-----PN---NPVFW-----SRDGGATWQAVPGATFG-F 240  
Db 243 LG-VTAVFPLPVQETQNDANDVVPNSDANQNVYMTENYFSPDRRYAVYKAAAGGPTAEF 301

```
Qy 241 IPHKGVPVNVHLVLIATSNVTGGPYDGGSDGVWKFSTGVTGWTTRISPVSPST----- 291
Db 302 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 348
Qy 292 -----DTANDYF-GYSLGTLIDRQHPNTI---MVATQISWPDFTIIFRSTGGGATWTRIW 341
Db 349 ATYYELTSGNOYFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 400
Qy 342 DWTS-YPNRSLRYVLDISAEPLWLTGVPQNPVPSPKLGWMDAEMAD-----PENSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVINRILREFTVRPAA 451
Qy 397 YGTGATLYATNDLTWKDGGQI-H-IAPMVKGLEE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA-----EPWAI GGSYQLGGPPQGWSEWNGLFRDSLRAQNEL-----GSMTI 502
Qy 446 SALGDLGGFTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVR-----AGS 493
Db 503 YVIQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVIYSCNGANNQAWP 556
Qy 494 FDPSSQPNDRHVAFTDGGK---NWFQSPGGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVQRRRAARTGMAFEMLSAGT 601
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAANAQIRSDR-----V 577
Db 602 PLMOGGBEYLRLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQLIAPRKAHPAL 658
Qy 578 NPKTFYALNSGTFRSTDDGVTGQVFPVAGLPSGAVG-----VMFHAVPGKEGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 700
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVGFKSAPG 665
Db 701 --LGDSNSIYVAYNGWSSSVTFLPAPPSTGQWYRVTDTCWMDGASTFV-----APG 751
Qy 666 SSYPAVFVVGITGGVGTAYRSDCGTTWVLNDHQHGYGNWQAI 710
Db 752 SE-----TLIGG-----AGTT-----YGCQGSQSL 770

RESULT 10
5457037-3
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1,797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO:3
; LENGTH: 750
5457037-3

Query Match 3.7%; Score 148.5; DB 6; Length 750;
Best Local Similarity 20.7%; Pred. No. 0.013;
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

Qy 27 GAPGILYVRTDIGMYRWDANGRWIPLLDVWGNWNGYNGWSIAADPINTKNVWAAVG 86
Db 70 GITGAVY-----YGYRWGPNWVPASWNGKSQAGFVSDVDANGDRFNPKNLLDDPY 121
Qy 87 MYTNSWDPNDGAILRSSDQAGATWQITPLPFKLGGMFGRGMGERLAVDPNNDNLIYFCAP 146
Db 122 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 156
```

```
Qy 147 SGKGLWRSTDSGATWSQMTNFPDV--GTVIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 157 KGVVLVPSTQSTGTPTKPTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 216
Qy 202 LQASKTTIFVGAD-----PN---NPVFM-----SRDGGATWQAVPGATG-F 240
Db 217 LG-VTAVBFLPVQETQNDANDVVPNSDANQNTYGYNTENYFSPDRRYAYNKAAGGPTAEP 275
Qy 241 IPHKGVPVNVHLVLIATSNVTGGPYDGGSDGVWKFSTGVTGWTTRISPVSPST----- 291
Db 276 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 322
Qy 292 -----DTANDYF-GYSLGTLIDRQHPNTI---MVATQISWPDFTIIFRSTGGGATWTRIW 341
Db 323 TTYELTSGNOYFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 374
Qy 342 DWTS-YPNRSLRYVLDISAEPLWLTGVPQNPVPSPKLGWMDAEMAD-----PENSDRML 396
Db 375 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDDAADSNAVINRILREFTVRPAA 425
Qy 397 YGTGATLYATNDLTWKDGGQI-H-IAPMVKGLEE-----TAVNDLISPPSGAPLI 445
Db 426 GSGGLDLFA-----EPWAI GGSYQLGGPPQGWSEWNGLFRDSLRAQNEL-----GSMTI 476
Qy 446 SALGDLGGFTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVR-----AGS 493
Db 477 YVIQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVIYSCNGANNQAWP 530
Qy 494 FDPSSQPNDRHVAFTDGGK---NWFQSPGGVTTGGTV---AASADGSRF-VWAPGD 545
Db 531 YGPS-----DGGTSTNYSWDQMSAG---TGAADVQRRRAARTGMAFEMLSAGT 575
Qy 546 P-----GQPVVYAVGFGN-----SWAASQGVPAANAQIRSDR-----V 577
Db 576 PLMOGGBEYLRLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQLIAPRKAHPAL 632
Qy 578 NPKTFYALNSGTFRSTDDGVTGQVFPVAGLPSGAVG-----VMFHAVPGKEGD 626
Db 633 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 674
Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVGFKSAPG 665
Db 675 --LGDSNSIYVAYNGWSSSVTFLPAPPSTGQWYRVTDTCWMDGASTFV-----APG 725
Qy 666 SSYPAVFVVGITGGVGTAYRSDCGTTWVLNDHQHGYGNWQAI 710
Db 726 SE-----TLIGG-----AGTT-----YGCQGSQSL 744

RESULT 11
5457037-5
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1,797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO:5
; LENGTH: 751
5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;
Best Local Similarity 20.7%; Pred. No. 0.013;
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;
```

```

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDWGVNWNWYNGVVSIAADPINTNKVAAVG 86
Db 71 GITGAVY-----YGYRWPENWPNYASNGKGSQAGFVSDVDANGDRFNPNKLLDDPY 122
QY 87 MYTNSWDPNDGAILRSSDOGATWQITPLPFKLGNNMGRGMBERLAVDPNNNDNIIYFCAP 146
Db 123 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157
QY 147 SKGGLWRSTDGATWSQMTNPPDV--GTYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 158 KGVLPVSTQSTGCTKPTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 217
QY 202 LGQASKTIIFGVAD-----PN--NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 218 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGPTAEF 276
QY 241 IPHKGVDPVNHVLYIATNTGPGYDSSGDVWKFVSTGWTTRISVPST-----291
Db 277 QAMVQAFNAGIKVYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 323
QY 292 -----DTANDYF-GYSGLTIDROHPNTI--MVATOISWMPDTIIFRSTDGATWTRIW 341
Db 324 TTYIELTSGNQFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM--GVDGFR-----F 375
QY 342 DWTs-YPNRSRLRYVLDISAEPMLTFGVQPNPPVPSKLGWMDMAID----PFSNRML 396
Db 376 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDAADSNVAINRILREFTVRPAA 426
QY 397 YGTGATLYATNDLTWKDSSGQI-H-IAPMVKGL-EE-----TAVNDLISPSPGAPLI 445
Db 427 GSGGLDLFA-----EPWAI GNSYQLGGFPQGWSEWNLFRDLSRQONEL-----GSMTI 477
QY 446 SALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINLPSIIVR-----AGS 493
Db 478 YVTQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMLTKDVYSCNANNQAWP 531
QY 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPBGVTTGGTV---AASADGSRF-VWAPGD 545
Db 532 YGPS-----DGGTSTNYSWDQMSAG---TGAADVORRAARTGMAFEMLSAGT 576
QY 546 P-----GQPVVYAVGFCN-----SWAASQGVPAANAQIRSDR-----V 577
Db 577 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQRLLIAFRKAHPAL 633
QY 578 NPKTFYALNSGTFFYRSTDGVTTFQPVAAAGLPSGAVG-----VMFHAVPGKEGD 626
Db 634 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYINGPS-- 675
QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSPAG 665
Db 676 --LGDSNSIYVAYNGWSSSVTFTLPAPSGTQWYRVTDTCWMDNGASTFV-----APG 726
QY 666 SSYPAVFVGTGTGGVTGAYRSDCGTTWVLINDDQHQYGNWGOAI 710
Db 727 SE-----TLIGG-----AGTT-----YGQCGQSL 745

```

## RESULT 12

```

US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

```

Query Match 3.7%; Score 148.5; DB 3; Length 776;

Best Local Similarity 20.7%; Pred. No. 0.013;

Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

```

QY 27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDWGVNWNWYNGVVSIAADPINTNKVAAVG 86
Db 96 GITGAVY-----YGYRWPENWPNYASNGKGSQAGFVSDVDANGDRFNPNKLLDDPY 147
QY 87 MYTNSWDPNDGAILRSSDOGATWQITPLPFKLGNNMGRGMBERLAVDPNNNDNIIYFCAP 146
Db 148 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
QY 147 SKGGLWRSTDGATWSQMTNPPDV--GTYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 183 KGVLPVSTQSTGCTKPTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 242
QY 202 LGQASKTIIFGVAD-----PN--NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGPTAEF 301
QY 241 IPHKGVDPVNHVLYIATNTGPGYDSSGDVWKFVSTGWTTRISVPST-----291
Db 302 QAMVQAFNAGIKVYMDV-----VYNHTAEGGTWTSDDPTTATIIYSWRGLDN 348
QY 292 -----DTANDYF-GYSGLTIDROHPNTI--MVATOISWMPDTIIFRSTDGATWTRIW 341
Db 349 TTYIELTSGNQFYDNTGIGANFNTYNTVAQNLI VDSLAYWANTM--GVDGFR-----F 400
QY 342 DWTs-YPNRSRLRYVLDISAEPMLTFGVQPNPPVPSKLGWMDMAID----PFSNRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDAADSNVAINRILREFTVRPAA 451
QY 397 YGTGATLYATNDLTWKDSSGQI-H-IAPMVKGL-EE-----TAVNDLISPSPGAPLI 445
Db 452 GSGGLDLFA-----EPWAI GNSYQLGGFPQGWSEWNLFRDLSRQONEL-----GSMTI 502
QY 446 SALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINLPSIIVR-----AGS 493
Db 503 YVTQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMLTKDVYSCNANNQAWP 556
QY 494 FDPSSQPNDRHVAFTDGGK-----NWFQSGPBGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVORRAARTGMAFEMLSAGT 601
QY 546 P-----GQPVVYAVGFCN-----SWAASQGVPAANAQIRSDR-----V 577
Db 602 PLMQGGDEYLRTLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFAQRLLIAFRKAHPAL 658
QY 578 NPKTFYALNSGTFFYRSTDGVTTFQPVAAAGLPSGAVG-----VMFHAVPGKEGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYINGPS-- 700
QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSPAG 665
Db 701 --LGDSNSIYVAYNGWSSSVTFTLPAPSGTQWYRVTDTCWMDNGASTFV-----APG 751
QY 666 SSYPAVFVGTGTGGVTGAYRSDCGTTWVLINDDQHQYGNWGOAI 710
Db 752 SE-----TLIGG-----AGTT-----YGQCGQSL 770

```

```

RESULT 13
US-09-341-461-2
; Sequence 2, Application US/09341461
; Patent No. 6586389
; GENERAL INFORMATION:
; APPLICANT: Hammond, Timothy G.
; APPLICANT: Verroust, Pierre J.
; TITLE OF INVENTION: Cubilin Protein, DNA Sequences Encoding Cubilin
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: D6148
; CURRENT APPLICATION NUMBER: US/09/341,461
; CURRENT FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: PCT/US99/01259
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 40
; SEQ ID NO 2
; LENGTH: 3623
; TYPE: PRT
; ORGANISM: rat
; FEATURE:
; OTHER INFORMATION: amino acid sequence of rat cubilin protein
US-09-341-461-2

Query Match
3.6%; Score 147; DB 4; Length 3623;
Best Local Similarity 20.0%; Pred. No. 0.14;
Matches 181; Conservative 79; Mismatches 301; Indels 344; Gaps 45;

Qy 6 YTWNSVAIGGGF-VDGIVNEGAGLTYRTDGGMYRWDAAANGRWIPLLDWGNWNG 64
Db 1731 YSASTSCGGSFYLDGIFNSDPY-----ADYG-----PNAECV-WN--- 1767
Qy 65 YNGVVSIAADP-----INTNK-----VMAAVGMYTNSWDNDGA 98
Db 1768 -----IASFGNRLQLSFLNLSNCKNDFVEIREGNATGGLIGRGCNSLPN-- 1819
Qy 99 ILRSSDQATWQITPLPKLGNMPPGMRGLAVDPNNNLYFGAPSGKGLWRST--- 155
Db 1820 -YSSAEGGSLW---VRFVSDGSGTGMGFQARFNKFNENNNIVGTGKIASPFWPKY 1874
Qy 156 DSGATW-----SMTN-FDP-----VGTIYANPTDTT 181
Db 1875 NSNYKVVNVNDAYGIIGRILEMDIEFTTNCFYDSLKIYDFGTGSLIGTYCG--TQTE 1932
Qy 182 GYSDIQGVVW-VAFDKSSSLGQASKTI FGVADPNPNPVFWSRDSGGATWQAVPGTGF 240
Db 1933 SFSSSRNLYTFQSSDSVSGRGLLEWPAVDVSDSTPPI-----APGACGGF 1981
Qy 241 I-----PHKGVDPVNHVLYIATSNITGPGYDGGSDVWKFVSTGTWTRISFPVSTDTAND 296
Db 1982 MVTGDTVPVGFSP-----GWPREVANG-----ADCIWIIYAP-DSTVELNI 2021
Qy 297 YFGYSGLTIDRQHPNTIMVATQISWPDPTIIFRSTGGATWTRIDWTSYPNRSLRVLD 356
Db 2022 -----LSLD-----IBPQSCNDKLVKDGDS-----D 2045
Qy 357 ISAEPMITFCVQNPVPSPKLGWMDAMADPNNSDRMLYGTG----- 400
Db 2046 LSPELAVLCGSPGPIRS-----TGEYMI-RTSDTSVAGTGFNASFGKSCGGYLAD 2099
Qy 401 -----ATLYATNDLTKWDSGGQIHAPMVKGLEE-----TAVNDLI---S 437
Db 2100 RGVTSPKYPTDYLPLNLCNSGVLVQGLTIIV-GFEQPFQIQNRDSFCQGDYLVLRNG 2158
Qy 438 PPSGAPLISALGDLGGTHADVTAVPSTIIFTSVFTTGTSDVYAEALNPSIIVRAGSPDPS 497
Db 2159 PDNGSPPLGSRNGRFGC---MYAPSTLFTS----- 2187
Qy 498 SQPNDRHAF---STDGKNWFGQSEGGVTTGTV-AAADGSRFWAPGDCPQPVYA 553
Db 2188 --GNEMFVQFISDSSNGGQFKIRYEAKSLACCGTVYIGDADSGYLTSPNYAN---YP 2242
Qy 554 VFGNSWAASQGVFANAIQIR-----SDRWNPXTFY-----ALSNGTFRSTDGGVTF 600

```

```

Db 2243 QGAEICWILEAPGCRSIOLEFEDQFQNIEDTPNCVSUYLELRDANSNARLVSKLCGGT-- 2300
Qy 601 QPVAAGLPSSGAVGMFHAVPKEGDMLAASSGLY--HSTNGSGSSWSAITGVSSAVNVG 658
Db 2301 -----LPGS-----WVSSRRIRYLFKGTGGSSYMGFKAKYSTASCG 2337
Qy 659 PGKAPGS-----SYPA-----VPVGTIGVTCAY-----RSDDCGTTWVLI 696
Db 2338 GTVSGSDGVIESIGYPTLFPYANNVFCQWFIRGLPGGYLTLSFEDFNLOSPPGCTKDFVEI 2397
Qy 697 NDD-----QHOXGNMGCAITGDHANLRRVYIGTG-----RGIVVGDIG 735
Db 2398 WENTSGRVLGRYCONSTPSSVDTSSNVASVXFVTDGSGVTASGFRLOFKSSRQVCGDLG 2457
Qy 736 GAPSG 740
Db 2458 G-PTG 2461

RESULT 14
US-09-640-419C-28
; Sequence 28, Application US/09640419C
; Patent No. 6630815
; GENERAL INFORMATION:
; APPLICANT: Bidney, Dennis L
; APPLICANT: Crasta, Oswald R
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE
; FILE REFERENCE: 35718/199009 (5718-92)
; CURRENT APPLICATION NUMBER: US/09/640,419C
; CURRENT FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,656
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/206,405
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1749
; TYPE: PRT
; ORGANISM: Synecocystis PCC6803
US-09-640-419C-28

Query Match
3.6%; Score 146; DB 4; Length 1749;
Best Local Similarity 21.8%; Pred. No. 0.062;
Matches 160; Conservative 81; Mismatches 306; Indels 188; Gaps 35;

Qy 119 GGN-----MPGRGM--GERLAVDPNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVG 171
Db 119 GGNANLFMNPAGMIFGNASINVPDGFVTTGSAIGFG-----NDQMFQVFSNDYN 171
Qy 172 TYIANPTDTTGYQSDIQGVVWVAFKSSSL-----GQASKT-----IFVGVADPNPNPVFW 222
Db 172 ALIGNPSQ-----FAFDLANPGLIINAGDLSVTEGKNLTLFAGNIVNTGSLA 218
Qy 223 SRDGGATWQAVPGTGFIPHKG-----VFDPNH-----VLYIATSNITGPDYD--- 266
Db 219 APGGNITVAAPGQNRIRISQAGSLLSLEVEVSPQMNQGGSPFVLDPLTLTQGSNLDL 278
Qy 267 ----GSSGDVWKFVSTGTWTRISVPVSTDTANDYFGYSGLTIDRQHPNTIMVA-TQISW 321
Db 279 GLAVQPNQSV-----TTNGTNALVSLPGSVTISGNVDASGKSTNLSGGQVVAIGQIAV 334
Qy 322 WPDITIFRSTDGGATWTRIDWTSYPNRSLRVLDISAEPMITFCVQNPVPSPKLGW 381
Db 335 QGATVDVSGNGGGGT-----VRIGGDFQGO--LT-----LPNASQTLI 370
Qy 382 DEAMADPFNSDRMLYGTGATLYA-TNDLTKWDS-----GGQIHAPMVKGLEET-- 430
Db 371 DSNVW---KADALLTGNGGTIVWADDSTRFSGNISAGGTMGGN-----GGFVETS 420
Qy 431 -AVNDLISPPSGAPLISALGDLGGF-----THADVTAVPSTIFTSPVTTGTSVD 479

```

Db 421 GAKSLMVDVDTARVNTFATMGBLGTWLLDPLBIIVGTDDLLADPKLV---SVLTITTSLD 477  
Qy 480 ---YAEINPGLIIRAG-SFDPSSOPN-----DRHVAFTDGGKKNWFOGSEPGVTT 526  
Db 478 NGNVILQADQSIQVAFSADPSAPGNLTFTSPITTDALSLGTGSIIFANTGP--INT 535  
Qy 527 GGTVAAS-----ADGSRFVWAPGDPGPVYVYVAGFNGSWAASQGVPPANAQI-- 572  
Db 536 GNTLVTSPTNLDPNKIQLNANTTFTAPG-----YDIYFRKSVNGGFDLLGNANFVY 588  
Qy 573 ---RSDRVNPKTFVSLNGTYRSTG---GGVTFQPVAAGLP-----SSGAVGMFHAVP 621  
Db 589 FDDGAGITPLPKSFVGTATEYIVGNNDIVTQGNQIFDGVFYGLQPVNLTSSAGSVFTNNI 648  
Qy 622 GKEGDLWLAASGLYHSTNGSSSAITGVSSAV-----NVGFGK-----SAPGS 666  
Db 649 LLNSGLQVQTAQIV--SQPSSLSAVEIADVLLNAGQNVSGNINTRGGNVDIQLGN 706  
Qy 667 SYPAVFVVGTTGGVGTAYRSDCGTTWVLYLNDQHQYGNWQCAITGDHANLRRVYIGTN- 725  
Db 707 ISTGSIVTSPFGNAGNVILNAGGT--LTTGVYETSGINGGDVTTSSGSNTSTAYIDTRG 764  
Qy 726 -GRGIVYGDIGCAPS 739  
Db 765 FGDGLEIDSLGGAVS 779

RESULT 15

US-08-362-525-22  
; Sequence 22, Application US/08362525  
; Patent No. 6027910  
; GENERAL INFORMATION:  
; APPLICANT: KLIS, FRANCISCUS M.  
; APPLICANT: SCHREUDER, MAARTEN P.  
; APPLICANT: TOSCHKA, HOLSER Y.  
; APPLICANT: VERRIPS, CORNELIS T.  
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE  
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,525  
; FILING DATE: 04-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92202080.5  
; FILING DATE: 08-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92203899.7  
; FILING DATE: 14-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/01763  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 213289/T7020 (V)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944

; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 894 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-525-22  
  
Query Match 3.5%; Score 143; DB 3; Length 894;  
Best Local Similarity 20.5%; Pred No. 0.041;  
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;  
  
Qy 7 TWSNVAIGGGFVDGIVFNEGAPGLIYVRTDTCGMWRDAANGRWIPLLDWVG----- 59  
Db 53 TVSNAAYMAYGA-----SKTKLGSVGGQTDISIDYN-----IPCSSSGSTFPQCPQE 99  
Qy 60 ---WNRWNGVGVVSIADPINTNKV---WAA---VGMVYNSWD----- 93  
Db 100 DSYGNWCKGKMGACS-----NSQIAYWSTDLFGFTTPTNVTLEMTGYELPQGTGSYTFK 155  
Qy 94 ---PNDGAILRSSDOGATW-----QITPLPKLGGNMPGRGMRGLAVDPNNDNL 141  
Db 156 FATVDDSAI---SVGGATAFNCCAOQPITSTNFTIDGIKPWG-----SLPNIEGTV 208  
Qy 142 YFGA---PSGKGLWRSTDGATWSQMTNFPDVGTVIANPTDTTGYQSDIQGVVWVAFDK 197  
Db 209 YWAGYVYFPM-KVYVSNVAVSWGTLPISVTLDP-GITVSD-----DFEGVYV-SFD- 255  
Qy 198 SSSSLGQASKTIFVGVADPNPNVFWSRDGGATQAVPGAPTGFPHKGVDPDPVNVHLYIA 257  
Db 256 ---DDLQSNCT---VPDPN-----YAVSTTTTTPWTGFTSTSTEMTIV 297  
Qy 258 TSNTGGPYD-----GSSGDMVKFSVTSGTWTRISVPVSTDTANDFYGYSGLTID-- 306  
Db 298 TGTNGVPTDETIVIVIRTPTEGLI---STTEPWTGFTSTSTEVTT-ITGNGQPTDET 353  
Qy 307 ---RQHPNTIMVATQISWPDITIFRSDGATWTRIDWTSYPNRSRLRYL----- 355  
Db 354 VIVIRTPTEGLISTTEPWTGFTSTSTE---MTVTGTNGQPTDETIVIVIRTPTEG 409  
Qy 356 ---DISAEPWL-TFGVQPNPVPSPKLGWMDAEMADPFNSDRMLYGTGATLYATNDLTKW 412  
Db 410 LVTTTTEPWTGTF-----TSTSTEMSTVTGNGLPTDETIV-- 444  
Qy 413 DSGGQIHIAPMVKGLEETAVNDLISPPSCA-----PLISALGDLGGFTHADVAV 462  
Db 445 -----IVKTFPTTAISSLSLSSSSSGQITSSITSRPIITPFYPNG-TSVISSV 493  
Qy 463 PSTIFTSPVFTTGTSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFTDGGKKNWFOGSEPG 522  
Db 494 ISSSVTSSLFTSPVSISSSVISSTTTSIIFESKSKSVIPTSSSTSSSE-SETSSAG 552  
Qy 523 GVTGTGTVAAADGSRFVWAPGDPGPVYVYVAGFNGSWAASQGVPPANAQIRSRVNPKT 582  
Db 553 SVSSSSFISSSESKSP---TYSSSSLPVTSATTSQETASSL-PPATTTKTSQTTLV- 607  
Qy 583 YALSNGTFYRSTDDGVTFOVPAAGLPSSGAVGMFHAVPG--KEGDLWLAAS----- 633  
Db 608 -----VTSCEHVCTESIPAVISTATV-----TVSGVTTEYTWCPISATTTKOT 654  
Qy 634 -GLYHSTNGSSWSAITGVSSAVNVYFGKSPGSSYPVAF--VVGITGGVTGAYRSDCG 690  
Db 655 KGTTEQTTETTKQTTVVTISSCESDVCSKTAS-----PAIVSISTATINGVTTEY----- 704  
Qy 691 TTWVLIN 697  
Db 705 TTWCPLIS 711



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 86.3468 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09E\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	12	US-09-917-376-3
2	4036	100.0	740	12	US-09-917-376-6
3	4036	100.0	740	14	US-10-155-400-3
4	4036	100.0	740	14	US-10-155-400-6
5	4036	100.0	957	12	US-09-917-376-1
6	4036	100.0	957	14	US-10-155-400-1
7	2429.5	60.2	882	14	US-10-156-761-9395
8	1680	41.6	726	12	US-09-917-376-7
9	1680	41.6	726	14	US-10-155-400-7
10	1579	39.1	838	12	US-10-420-191-2
11	1573.5	39.0	818	14	US-10-026-994-2
12	1405	34.8	739	14	US-10-156-761-10111
13	1117	27.7	789	12	US-10-395-241-14
14	1117	27.7	812	12	US-10-395-241-12
15	1114	27.6	826	12	US-10-395-241-18

16	841.5	20.8	555	10	US-09-927-827-47	Sequence 47, Appl
17	238	5.9	2468	12	US-10-282-122A-66335	Sequence 66335, A
18	238	5.9	2468	12	US-10-246-330-4	Sequence 4, Appl
19	199.5	4.9	2435	12	US-10-282-122A-47453	Sequence 47453, A
20	193	4.8	1465	15	US-10-369-493-13955	Sequence 13955, A
21	181	4.5	1308	12	US-10-282-122A-44566	Sequence 44566, A
22	172.5	4.3	783	15	US-10-369-493-14105	Sequence 14105, A
23	170	4.2	1074	12	US-10-282-122A-50816	Sequence 50816, A
24	168.5	4.2	1439	12	US-10-282-122A-44567	Sequence 44567, A
25	167.5	4.2	1119	14	US-10-245-802-12	Sequence 12, Appl
26	165.5	4.1	613	15	US-10-369-493-11181	Sequence 11181, A
27	163.5	4.1	1289	12	US-10-282-122A-44901	Sequence 44901, A
28	158.5	3.9	774	9	US-09-833-435A-8	Sequence 8, Appl
29	158.5	3.9	774	14	US-10-375-720-8	Sequence 8, Appl
30	157	3.9	350	14	US-10-156-761-9225	Sequence 9225, Ap
31	156.5	3.9	2204	12	US-10-282-122A-64364	Sequence 64364, A
32	156	3.9	2117	15	US-10-130-801-63	Sequence 63, Appl
33	152	3.8	827	12	US-10-282-122A-47312	Sequence 47312, A
34	152	3.8	1721	12	US-10-282-122A-62548	Sequence 62548, A
35	152	3.8	3472	13	US-10-027-806-4	Sequence 4, Appl
36	152	3.8	3472	13	US-10-034-623-4	Sequence 4, Appl
37	152	3.8	3472	14	US-10-027-801-4	Sequence 4, Appl
38	152	3.8	3472	14	US-10-029-120-4	Sequence 4, Appl
39	151.5	3.8	776	9	US-09-833-435A-4	Sequence 4, Appl
40	151.5	3.8	776	14	US-10-375-720-4	Sequence 4, Appl
41	151	3.7	1291	14	US-10-156-761-14161	Sequence 14161, A
42	151	3.7	1325	12	US-10-282-122A-44953	Sequence 44953, A
43	151	3.7	1797	15	US-10-369-493-5176	Sequence 5176, Ap
44	151	3.7	1805	15	US-10-369-493-5177	Sequence 5177, Ap
45	150.5	3.7	1082	15	US-10-369-493-17890	Sequence 17890, A

#### ALIGNMENTS

#### RESULT 1

US-09-917-376-3  
; Sequence 3, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-09-917-376-3

Query Match	100.0%;	Score 4036;	DB 12;	Length 740;
Best Local Similarity	100.0%;	Pred. No. 2.2e-301;		
Matches 740;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDTCGMTRWDANGRWIPLDVGW	60	
Db	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDTCGMTRWDANGRWIPLDVGW	60	
Qy	61	NNWGVGVSTAADPINTNKWAAVGMVTSNDNDGALLRSSDQCATQWITLPPFKLGG	120	
Db	61	NNWGVGVSTAADPINTNKWAAVGMVTSNDNDGALLRSSDQCATQWITLPPFKLGG	120	
Qy	121	NNPGRGMGERLAVDPNNDNIFYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT	180	

Db 121 NMPGRGMRERLAVDPNNDNLIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180  
 Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGADPNPNPFWSRDGGATWQAVPGAPTGF 240  
 Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGADPNPNPFWSRDGGATWQAVPGAPTGF 240  
 Qy 241 IPHGKGVDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSITDANDYFGY 300  
 Db 241 IPHGKGVDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSITDANDYFGY 300  
 Qy 301 SGLTIDRQHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
 Db 301 SGLTIDRQHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
 Qy 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
 Db 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
 Qy 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
 Db 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
 Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQGVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDDGVT 600  
 Db 541 WAPGDPQGVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDDGVT 600  
 Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSPAGSSYPAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRV 720  
 Db 661 KSPAGSSYPAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRV 720  
 Qy 721 YIGTNGRGIVYGDIGGAPSG 740  
 Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 2

US-09-917-376-6  
 ; Sequence 6, Application US/09917376  
 ; Publication No. US20040038334A1  
 ; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYLICUS  
 ; FILE REFERENCE: 40197.4US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,376  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus

US-09-917-376-6  
 Query Match 100.0%; Score 4036; DB 12; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-301;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPYTWSNVAICGGFVDGIVNEGAPGLIYRTDIDGMYRWDAANGRWIPLLDWVGW 60  
 Db 1 ATTQPYTWSNVAICGGFVDGIVNEGAPGLIYRTDIDGMYRWDAANGRWIPLLDWVGW 60

Qy 61 NNWGTNGVVSIAADPINTNKVWAAVGMYTNSWDPNDGAILRSSDQGAQWITPLPFLKGG 120  
 Db 61 NNWGTNGVVSIAADPINTNKVWAAVGMYTNSWDPNDGAILRSSDQGAQWITPLPFLKGG 120  
 Qy 121 NMPGRGMRERLAVDPNNDNLIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180  
 Db 121 NMPGRGMRERLAVDPNNDNLIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180  
 Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGADPNPNPFWSRDGGATWQAVPGAPTGF 240  
 Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGADPNPNPFWSRDGGATWQAVPGAPTGF 240  
 Qy 241 IPHGKGVDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSITDANDYFGY 300  
 Db 241 IPHGKGVDPVNHVLYIATNTGGPYDGSGLWKFVSCTGTWTRISPVPSITDANDYFGY 300  
 Qy 301 SGLTIDRQHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
 Db 301 SGLTIDRQHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSYPNRSIRYVLDISAE 360  
 Qy 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
 Db 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
 Qy 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
 Db 421 APWKGLBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
 Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIVRAGSDPSSQPNDRHVAFSTDDGKNWFGSGEGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQGVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDDGVT 600  
 Db 541 WAPGDPQGVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDDGVT 600  
 Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSPAGSSYPAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRV 720  
 Db 661 KSPAGSSYPAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRV 720  
 Qy 721 YIGTNGRGIVYGDIGGAPSG 740  
 Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 3

US-10-155-400-3  
 ; Sequence 3, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYLICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE:  
 ; OTHER INFORMATION: Catalytic domain GH74

US-10-155-400-3

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 2.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-10-155-400-6

QY 1 ATTPYTWNSNAVAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 1 ATTPYTWNSNAVAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
QY 61 NNWYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNNPVFWSRDGATWQAVPGATGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNNPVFWSRDGATWQAVPGATGF 240  
QY 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
QY 361 PWTFTGVQPNPVPSPKLGWDEAMAI DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVQPNPVPSPKLGWDEAMAI DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APYKGLTEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSPVFTTGTSDY 480  
DB 421 APYKGLTEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSPVFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFFYRSTDCGVTF 600  
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFFYRSTDCGVTF 600  
QY 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
DB 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYDGGAPSG 740  
DB 721 YIGTNGRGIVYDGGAPSG 740

## RESULT 4

US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155.400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1

## RESULT 5

US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 2.2e-301;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-10-155-400-6

QY 1 ATTPYTWNSNAVAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 1 ATTPYTWNSNAVAGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
QY 61 NNWYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWYGVVSIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNNPVFWSRDGATWQAVPGATGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNNPVFWSRDGATWQAVPGATGF 240  
QY 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDSSGVDWKFVSVTGWTWTRISPVPSDTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPYNSRLRYVLDISAE 360  
QY 361 PWTFTGVQPNPVPSPKLGWDEAMAI DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVQPNPVPSPKLGWDEAMAI DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APYKGLTEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSPVFTTGTSDY 480  
DB 421 APYKGLTEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSPVFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGVTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFFYRSTDCGVTF 600  
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFFYRSTDCGVTF 600  
QY 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
DB 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQQAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYDGGAPSG 740  
DB 721 YIGTNGRGIVYDGGAPSG 740

; TITLE OF INVENTION: CELLULOYTICUS  
 ; FILE REFERENCE: 40197.4US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,376  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 957  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (957)  
 ; OTHER INFORMATION: Any amino acid  
 US-09-917-376-1

Query Match 100.0%; Score 4036; DB 12; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-301;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 DB 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
 QY 61 NNWNGVGVVSIADPINTNKWAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 120  
 DB 107 NNWNGVGVVSIADPINTNKWAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 166  
 QY 121 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
 DB 167 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 226  
 QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDPPNPFVMSRDGGATWQAVPGAPTGF 240  
 DB 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDPPNPFVMSRDGGATWQAVPGAPTGF 286  
 QY 241 IPHKGVPFVNVHLVYIATNTGCGPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY 300  
 DB 287 IPHKGVPFVNVHLVYIATNTGCGPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY 346  
 QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIIPRSTDGGATWTRIDWTSYPNRSRLRVLDISAE 360  
 DB 347 SGLTIDRQHPNTIMVATQISWMPDTIIIPRSTDGGATWTRIDWTSYPNRSRLRVLDISAE 406  
 QY 361 PMLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420  
 DB 407 PMLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 466  
 QY 421 APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 480  
 DB 467 APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 526  
 QY 481 AELNPSIIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV 540  
 DB 527 AELNPSIIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV 586  
 QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF 600  
 DB 587 WAPGDPGPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF 646  
 QY 601 QPVAAGLPSSGAVGVNMFHAPVKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 DB 647 QPVAAGLPSSGAVGVNMFHAPVKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
 QY 661 KSAFGSSYPVAVVGTIGVTVGAYRSDCCGTWVLINDDQHQYGNWGAITGDHANLRV 720  
 DB 707 KSAFGSSYPVAVVGTIGVTVGAYRSDCCGTWVLINDDQHQYGNWGAITGDHANLRV 766  
 QY 721 YIGTNGRGIVYGDIGGAPSG 740  
 DB 767 YIGTNGRGIVYGDIGGAPSG 786

RESULT 6  
 US-10-155-400-1  
 ; Sequence 1, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOYTICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 957  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (957)  
 ; OTHER INFORMATION: Any amino acid  
 US-10-155-400-1

Query Match 100.0%; Score 4036; DB 14; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-301;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 DB 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
 QY 61 NNWNGVGVVSIADPINTNKWAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 120  
 DB 107 NNWNGVGVVSIADPINTNKWAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 166  
 QY 121 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
 DB 167 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 226  
 QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDPPNPFVMSRDGGATWQAVPGAPTGF 240  
 DB 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDPPNPFVMSRDGGATWQAVPGAPTGF 286  
 QY 241 IPHKGVPFVNVHLVYIATNTGCGPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY 300  
 DB 287 IPHKGVPFVNVHLVYIATNTGCGPYDGGSDGVKFSVTSQGTWTRISPVSTDTANDYFGY 346  
 QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIIPRSTDGGATWTRIDWTSYPNRSRLRVLDISAE 360  
 DB 347 SGLTIDRQHPNTIMVATQISWMPDTIIIPRSTDGGATWTRIDWTSYPNRSRLRVLDISAE 406  
 QY 361 PMLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420  
 DB 407 PMLTFGVQPNPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 466  
 QY 421 APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 480  
 DB 467 APMVKGLBETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 526  
 QY 481 AELNPSIIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV 540  
 DB 527 AELNPSIIIVRAGSPDPSSQPNDRHVAFSTDGKWNFGQSEPGGVTTGGTVAASADGSRFV 586  
 QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF 600  
 DB 587 WAPGDPGPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFFRSTDCGVTF 646  
 QY 601 QPVAAGLPSSGAVGVNMFHAPVKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 DB 647 QPVAAGLPSSGAVGVNMFHAPVKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706



Db 352 DISNAPWLODTTSTDOFPV---RVGMVVEALALDPFDSNHMLYGTGLTVYCGHDLTNWDS 408  
QY 415 GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 474  
Db 409 KHNVTVKSLAVGIEEMAVLGLITPPGPGALLSAGVGGDFYHSDLDAAAPNQAYHTPYGT 468  
QY 475 GTSVDYAEINPSIIIVRAGSFPDSSOPNDRHVAFSTGDKNWFQSGEPGGVTTGGTVAASA 534  
Db 469 TNGIDYAGNKPNSIIVRSAGSD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 523  
QY 535 DGSRRVWAPGDPQPVVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFFYRST 594  
Db 524 DGDVTLLMSSSTGALVSKSQG---TLTAVSSLPSCAVIASDKSDNTVYFGSAGAIYVSK 580  
QY 595 DGVVTFQPVAAAGLPSSGAVGVMFHVPKGEGLWLAASSGLYHSTNGGSSWSAI-TGVSS 653  
Db 581 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638  
QY 654 AVNVGFGKSAPGSSYPANFVVGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWQA--IT 711  
Db 639 GWSFGFGKASSTGSGYVVIYGFFTIDGAAGLFKSEDAGTNWQVSDASHGFGS-GSANVNV 697  
QY 712 GDHANLRVYIGTNGRGIVYGDIGGAPSG 740  
Db 698 GDLQTYGRVFRGHERPHLLRSQREPAG 726  
RESULT 9  
US-10-155-400-7  
; Sequence 7, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155, 400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Aspergillus aculeatus  
US-10-155-400-7

Query Match 41.6%; Score 1680; DB 14; Length 726;  
Best Local Similarity 46.2%; Pred. No. 2.7e-120;  
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;  
QY 1 ATTQPTWNSVNAI-GGGGFGVIGNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59  
Db 1 AASQATWKNVVTGGGGGFTPGIVNPSAKGVAYARTDIGGAYRLN-SDDTWTPLMDWVG 59  
QY 60 ---NNWNYGVVSTAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQAGTQITPLPF 116  
Db 60 NDTWHDW---GIDALATPEVDTRVYVAVGMVYNEWDNPSGILRSTDQGTWTETKLPP 116  
QY 117 KLGNNMPGRGGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 176  
Db 117 KVGNNMPGRGGERLAVDPNNKNSILYFGARSGHGLWKTSDYGATWSNVTSTWTGYTQD 176  
QY 177 PTDFTGYQSDIQGVVWVAFDSSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPCA 236  
Db 177 SSST--YTSDPVGIAMVTFDSTSGSSGATPRIFVGVADAGKSVFKSEDAGATWAWVSGE 234  
QY 237 PT-GFIPHKGFDPVNNVLYIATSTNGGPDYSSGDVWKFVSTGTTWTRISPVPSDTDTAN 295  
Db 235 PQYGLPHKGLVSPEKTLTYISYANGAGPYDGTNGTHVKNITSGVMTDISP---TSLAS 291

QY 296 DYFGVSGLTIDRQHPNTIMVATQISWMPDIIIPRSTDGGATWTRINDWTSYPNRSIYVL 355  
Db 292 TYGYGGLSVLDQVPGTLMVAALNCWMPDELIPRSTDSGATWSPIMEWNGYPSINYYSY 351  
QY 356 DISASBPWLTFFGVQPNP-PVPSPKLGWMDRAMAIDPENSRLMYGTGATLVATNDLTKWDS 414  
Db 352 DISNAPWLODTTSTDOFPV---RVGMVVEALALDPFDSNHMLYGTGLTVYCGHDLTNWDS 408  
QY 415 GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 474  
Db 409 KHNVTVKSLAVGIEEMAVLGLITPPGPGALLSAGVGGDFYHSDLDAAAPNQAYHTPYGT 468  
QY 475 GTSVDYAEINPSIIIVRAGSFPDSSOPNDRHVAFSTGDKNWFQSGEPGGVTTGGTVAASA 534  
Db 469 TNGIDYAGNKPNSIIVRSAGSD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 523  
QY 535 DGSRRVWAPGDPQPVVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFFYRST 594  
Db 524 DGDVTLLMSSSTGALVSKSQG---TLTAVSSLPSCAVIASDKSDNTVYFGSAGAIYVSK 580  
QY 595 DGVVTFQPVAAAGLPSSGAVGVMFHVPKGEGLWLAASSGLYHSTNGGSSWSAI-TGVSS 653  
Db 581 NTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638  
QY 654 AVNVGFGKSAPGSSYPANFVVGTTGVTGAYRSDDCGTTWVLINDDQHOYGNWQA--IT 711  
Db 639 GWSFGFGKASSTGSGYVVIYGFFTIDGAAGLFKSEDAGTNWQVSDASHGFGS-GSANVNV 697  
QY 712 GDHANLRVYIGTNGRGIVYGDIGGAPSG 740  
Db 698 GDLQTYGRVFRGHERPHLLRSQREPAG 726  
RESULT 10  
US-10-420-191-2  
; Sequence 2, Application US/10420191  
; Publication No. US20040067569A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; FILE REFERENCE: 10210-200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; CURRENT FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; PRIOR FILING DATE: 2002-04-19  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent In version 3.2  
; SEQ ID NO 2  
; LENGTH: 838  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
US-10-420-191-2

Query Match 39.1%; Score 1579; DB 12; Length 838;  
Best Local Similarity 43.0%; Pred. No. 1.9e-112;  
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;  
QY 6 YTWSNVAI-GGGGFGVIGNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWV---GW 60  
Db 21 FSKVKNVKGGGGFGVPGIIFHFKTKGVAYARTDIGGLYRLN-ADDSWTAVTDGIADNAGW 79  
QY 61 NNWYGVVYIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQAGTQITPLPKLGG 120  
Db 80 HNW---GIDAVALDQDDQKVYAAVGMVYNSWDPNSGAILRSSDRCATWSFTNLPKVGG 136  
QY 121 NMPGRGGERLAVDPNNINILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTD 180  
Db 137 NMPGRGGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFKSVSFTATGTYPDPSDS 196

Qy 181 TGYOSDIQGVVWVAFDKSSSLGQASKTIFVGVADP--NNPVFWRDGGATWQAVPGAPTG 239  
 Db 197 NGYNSDKQGLMWVTFDSTSTTGCATSRIFGTADNITASVYVSTNAGSTWSAIVPGPGK 256  
 Qy 240 FIPKGVDPVNVHVLXIATNTGTPYDGGSSGVKESVTSWTWTRISPVSTDTANDYFG 299  
 Db 257 YFPHAKLOPAEKALYITYSDGTGPDGLGVSVMRYDIAGTWKDIITPVSGSDL---YFG 313  
 Qy 300 YSGLTIDRQHNTIMVATQISWNPDTIIFRSTGGATWTRIDWTSYPNRSLRYVLDISA 359  
 Db 314 FGLGLDLQKPGTLVVASLNSWPDQALFRSTDGTTWSPINAWASYPTETYYYSISTPK 373  
 Qy 360 EPWLTFG--VQNPVPVPS----PKLGMDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDS 414  
 Db 374 APWIKNNFIDVTSBSPDGLIKRLGWMIESLEIDPTDSNHWLYGTGWTIFGCHDLTNWDT 433  
 Qy 415 GGQIHAPMKVGLBETAUNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471  
 Db 434 RHNVSIQSLADGIEEFSVQDLASAPGSELLAAVGDNDNGFTFASRNDLGTSPQTVWATPT 493  
 Qy 472 FTTGTSVDYAEINPSIIVRAGSPDPSSQPNDRHVAFTDGGKNWFOGSEPGVTTGGTVA 531  
 Db 494 WATSTVDYAGNSVKSVVRVGN-----VAISDGGATWSDIYAADTSMNGGTVA 548  
 Qy 532 ASADGSRFVWAPGDPGPVYVAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFY 591  
 Db 549 YSADGDTILWSTASSG---VQRSQFGSFAVSLLPAGAVIASDKKTNVSFYAGSGSTFY 605  
 Qy 592 RSTGGVTFQVVAAGLPSGAVGVW--FHAVPGKEGDLWLAASGLYHSTNGGSSWSAI- 648  
 Db 606 VSKDTGSSF---TRG-PKLSAGTIRIDAAHPTTAGTLYVSTDVGIIFRSTDGTTFGQVS 661  
 Qy 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVGTGA--YRSDCGTWTWVLINDDQHQYGNW 706  
 Db 662 TALNTYQIALGVGS--GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQGSGFGSID 717  
 Qy 707 QQAITGDHANLRVYVGTNGRGIVY--GDIGGAPSG 740  
 Db 718 STKVAGSGSTAGQVYVGTNGRGVFAOQTVGGTGG 753

RESULT 11  
 US-10-026-994-2  
 ; Sequence 2, Application US/10026994  
 ; Publication No. US20030113732A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Goedegebuur, Frits  
 ; APPLICANT: Ward, Michael  
 ; APPLICANT: Yao, Jian  
 ; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids  
 ; TITLE OF INVENTION: Encoding the Same  
 ; FILE REFERENCE: GC698  
 ; CURRENT APPLICATION NUMBER: US/10/026,994  
 ; CURRENT FILING DATE: 2002-04-30  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 818  
 ; TYPE: PRT  
 ; ORGANISM: Trichoderma reesei  
 US-10-026-994-2

Query Match 39.0%; Score 1573.5; DB 14; Length 818;  
 Best Local Similarity 42.9%; Pred. No. 4.9e-112;  
 Matches 324; Conservative 122; Mismatches 265; Indels 45; Gaps 20;

Qy 6 YTWSNVAI-GGGGVDGIVFNEGAPGLYVTRDTIGMYRWDAANGRWIPLLDWV---GW 60  
 Db 2 FSWKNVILGGGGVPGIIFPHKTKGVAYARTDGGYRLN-ADDSWATVTDGADNAGW 60  
 Qy 61 NNWGYNGVVSIAADPINTNKVAAVGMVYTNWSDNDGAILRSSDQGWATQITPLPFKLG 120

Db 61 HNW---GIDAVALDDQDKVVAVGMVTNSWDPNSGAILRSSDRGATWSFTNLPRKVG 117  
 Qy 121 NMPGRGEMGERLAVDPNNDNIIIFYGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180  
 Db 118 NMPGRGAGERLAVDPANSNIIIFYGARSNGLWKSTDGGVTFKSVSFTATGTVIPDPSDS 177  
 Qy 181 TGYOSDIQGVVWVAFDKSSSLGQASKTIFVGVADP--NNPVFWRDGGATWQAVPGAPTG 239  
 Db 178 NGYNSDKQGLMWVTFDSTSTTGCATSRIFGTADNITASVYVSTNAGSTWSAIVPGPGK 237  
 Qy 240 FIPKGVDPVNVHVLXIATNTGTPYDGGSSGVKESVTSWTWTRISPVSTDTANDYFG 299  
 Db 238 YFPHAKLOPAEKALYITYSDGTGPDGLGVSVMRYDIAGTWKDIITPVSGSDL---YFG 294  
 Qy 300 YSGLTIDRQHNTIMVATQISWNPDTIIFRSTGGATWTRIDWTSYPNRSLRYVLDISA 359  
 Db 295 FGLGLDLQKPGTLVVASLNSWPDQALFRSTDGTTWSPINAWASYPTETYYYSISTPK 354  
 Qy 360 EPWLTFG--VQNPVPVPS----PKLGMDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDS 414  
 Db 355 APWIKNNFIDVTSBSPDGLIKRLGWMIESLEIDPTDSNHWLYGTGWTIFGCHDLTNWDT 414  
 Qy 415 GGQIHAPMKVGLBETAUNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471  
 Db 415 RHNVSIQSLADGIEEFSVQDLASAPGSELLAAVGDNDNGFTFASRNDLGTSPQTVWATPT 474  
 Qy 472 FTTGTSVDYAEINPSIIVRAGSPDPSSQPNDRHVAFTDGGKNWFOGSEPGVTTGGTVA 531  
 Db 475 WATSTVDYAGNSVKSVVRVGN-----TAGTQVAILSSDGGATWSDIYAADTSMNGGTVA 528  
 Qy 532 ASADGSRFVWAPGDPGPVYVAVFGNSWAASQGVPPANAQIRSDRVNPKTFYALSNGTFY 591  
 Db 529 YSADGDTILWSTASSG---VQRSQFGSFAVSLLPAGAVIASDKKTNVSFYAGSGSTFY 585  
 Qy 592 RSTGGVTFQVVAAGLPSGAVGVW--FHAVPGKEGDLWLAASGLYHSTNGGSSWSAI- 648  
 Db 586 VSKDTGSSF---TRG-PKLSAGTIRIDAAHPTTAGTLYVSTDVGIIFRSTDGTTFGQVS 641  
 Qy 649 TGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVGTGA--YRSDCGTWTWVLINDDQHQYGNW 706  
 Db 642 TALNTYQIALGVGS--GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQGSGFGSID 697  
 Qy 707 QQAITGDHANLRVYVGTNGRGIVY--GDIGGAPSG 740  
 Db 698 STKVAGSGSTAGQVYVGTNGRGVFAOQTVGGTGG 733

RESULT 12  
 US-10-156-761-10111  
 ; Sequence 10111, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 10111  
 ; LENGTH: 739  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-10111



```
Query Match      34.8%; Score 1405; DB 14; Length 739;
Best Local Similarity 40.8%; Pred. No. 3.8e-99;
Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;

QY 6 YTNVAIGGGFVDGIVFNEGAPGILYVRTDGGMYRWDAAANGRWIPLLDWVGWNNWGY 65
DB 42 YRNRNAVIGGTFVGVGLFHPFSVRGLAYARTDIGGAYWDDRGARWTLPLIDHLGWDWNL 101
QY 66 NGVWSIAADPINTNKVMAAVGMYTNSWDPNDGAILRSSDOGATWQITPLPEKLGNNMGR 125
DB 102 LGVEMAVDPHPDLYLAVGTAGSWAGN-GAVLRSEDGATWTRTDLTVKLGNEGR 160
QY 126 GNGERLAVDPNNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDITGYQ 185
DB 161 GAGERLLVDPDSDTLWLT-RHDGLLKSTDKGATAAAATAPP-----AKANS 207
QY 186 DIQGVVWVAFPKSSSLGQASKTIFVGVADPNP-----FWSRGGATWQAVCAPGPF 240
DB 208 SGQGVWF-----LVAAGRTVYAGWGDGSGTGTANLYRTAD-GTTWGAVPGRPSGT 257
QY 241 --IPHKGVFDPVNVHLYIATSNCGPYDGGSDGVWKFVSFTGTWTRISPV----PSTDT 293
DB 258 SAKVPLRAAYTHRELVTYGDAPCGQSDGSHKLRITAIGTWETVTPVKPGTTSDG 317
QY 294 ANDYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTGGATWTRIIDWTSYPNRSLRY 353
DB 318 SADTFAYGVAVDARRPOTLVVSTNNRWADGTVFRSDGRTWTSKX-----AA 368
QY 354 VLDLSAEPWLTFGVQPNPVPSPKLGWDEAMADPPNSDRMLYGTGATLYATNDLTW 413
DB 369 VFDVSETPFLDWGD-----KPKFGMWIOALAVDPYDSQHVYVGTGATLYTRDLKEW- 421
QY 414 SGGQIHIAPMVKGLEETAUNDLISPPSG-APLISALGDLGGFTHADVTAVPST-IFTSPV 471
DB 422 -----APRIRGLESAVRQLISIPVGEAHLISGLDGIWYHERLTASPSRGATNPV 474
QY 472 FTTGTSVDYAEINPSIIVRAGSFDPSSQPNDRHVAFASTDDGKNW--FGSEPPGVTTGGT 529
DB 475 FGSATGLAQAAARPAYVVRTGWDHNG-----AYSHDGGRTWAPFEAQPIAKDAPG 528
QY 530 VAASADGSRFWA-----PGDGPQPVVYA-----VGFNGSWAASQGVPAQAQIRSRVNPKT 581
DB 529 IATSDAGGTLTWSFVHDG-----TTAAHRSTONGASWSEVSPKPKATPVADPADPTR 583
QY 582 FYA--LNGTYRSTGCGVTFQPVAAAGLPSSGAVGMFHAVPGEGDLWLAAS-SGLYHS 638
DB 584 FYADFNGTLTASTDSGRSFTARAGLP-SGDSQFKLVAAPGRSGDLWLSAKMNGLYRS 642
QY 639 TNGGSSWASITGVSSAVNVGFKSAPSSYPAVFVVGVTIGGVTGAYRSDDCGTTWVLND 698
DB 643 TDGGDTFARIDSCWASYTLFGKAADGADYPAIQVGSFTEITAVYRSDDAARTWVRIND 702
QY 699 DOHQYGNWQAITGDHANLRVYIGTNGRGIVYGD 733
DB 703 DAHQWGWIGEAVVGDPIRHGRVYLATNGRGIQYGE 737

RESULT 13
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; TITLE OF INVENTION: NOVEL XYLOGUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR APPLICATION NUMBER: 2003-03-25
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
; US-10-395-241-14

Query Match      27.7%; Score 1117; DB 12; Length 789;
Best Local Similarity 35.0%; Pred. No. 5.8e-77;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

QY 6 YTNVAIGGGFVDGIVFNEGAPGILYVRTDGGMYRWDAAANGRWIPLLDWVGWNNWGY 65
DB 4 YEFKVAIGGGYITGIVAHPKTKDLYARTDIGGAYRWDAGTSKMWPLNDFIEAQDMNI 63
QY 66 NGVWSIAADPINTNKVMAAVGMYT-NSWDPNDGAILRSSDOGATWQITPLPEKLGNNMGP 124
DB 64 MGTESIALDPNDPDRDLYLAQGRYVGDEW-----AAFVSEDRGQSFTIYESFPFMGANDMG 119
QY 125 RGMGERLAVDPNNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDITGYQ 184
DB 120 RNNGERLAVNPNSNEVWNGTRT-EGIKSSDRAKTWTNVTSPIDAF-----TNGIGYT 172
QY 185 SDIQGVVWVAFPKSSSLGQASKTIFVGVADPNPFWSRDGGATWQAVCAPGTFI--- 241
DB 173 S-----VIFDP-----ERNGTIYASATAPQG-MYVTHDGGVSEVPVAGQPSSSLNRT 218
QY 242 -----PHKGVFDPVNVHLYIATSNCGPYDGGSDGVWKFVSFTGTWTRII- 285
DB 219 TGAFPDKKPASTAPQPMKVALTP--NFLVTVYADYFGPVGWTVFGEVNRQNRSGAMDDIT 276
QY 286 -----SPVPSTDTANDYFGYGLTIDRQHPNTIMVATQISWMPDTI---IFRSTDCGAT 336
DB 277 PRVGNSSPAPYNNQIFPAGGFCGLSDVATNPRLVIT-LDRDPGALDSIYLSTDAGAT 335
QY 337 WTRINDWTS-----YPNRSRLRVLDISABPWITFGVQP-----NPVPSP---KLG 379
DB 336 WKDVTQLSSPSNLEGNWGHPTNAARY-KDGTVPVWLDFFNNGPQWGGYGAPHGTPGLTKFG 394
QY 380 WDEAMADPPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAP---MVKGLEETA 431
DB 395 WMSAVLIDPFPPEHLMYGTGATIWATDLSRVEKDW-----APSWYLQIDGIEENA 446
QY 432 VNDLISPPSCAPLISALGDLGGFTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVRA 491
DB 447 ILSLSRSPKSCAALLSGIGDISKMKHDDLTK-PQKMFGAPOFSNLDSDIDAAGNPNVVRRA 505
QY 492 GSFDPSQPNDRHVAFASTDDGKNW--FGSEPPGVTT--GGTVAASADGSRFWAP--G 544
DB 506 GSSGHEYDSACARGAYATDGDATWITFTPCPPGMNASHYQGSTIAVDASGSIWSTKLD 565
QY 545 DPGQPVVYAVGRGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFFRSTDDGV 598
DB 566 EQASGFWYSHDYKTKWS-----VPAGDLKAQTANVLSDKVQDGTFFYATDGGKFFVSTDDGK 621
QY 599 TPOVAAAGLPSSGAVGMFHAVPGEGDLWL-AAASGLYHSTNGGSSWSAI-TGVSSAVN 656
DB 622 SYAAKAGALVT--GTSLMPAVNPWVAGDVWVPVEGGLFHSTDFGASFTVGTGANATLVS 679
QY 657 VGFGKS-----APGSSYPVAVFVVGTT--IGGVTGAYRSDDCGTTWVLINDDOHQYGNWQAI 710
DB 680 VGAPKSKSGKKASAPSASFVITWTDKPGSDIGLYRSDNGSTWTRVNDQEHNTSG-PTMI 738
QY 711 TGDHANLRVYIGTNGRGIVYGD 734
DB 739 EADPKVYGRVLTNGRGIVYADL 762

RESULT 14
US-10-395-241-12
; Sequence 12, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR APPLICATION NUMBER: 2003-03-25
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
```

DD 507 GSSGHEYDSACARGAYATDGGDAWITFTPTCPGGMNASHIQGSI1AIVDASGSQILVWSKJID 566

Qy	545	DPGPPVYAVGFCGNSWAASQVPA-----NAQIRSDRVNPKTFYALSNGTFRSTDCGV	598
Db	567	EQASGPMYSHDYGKTWS-----VPAGDLKAQTANVLSDKVQDGTFYATDGGKFPVSTDCGK	622
Qy	599	TFQPVAAGLPSSGAVGVNMFHAVPGKEGDLWL-AASSGLYHSTNGGSSWSAI-TGVSSAVN	656
Db	623	SYAAKAGLVT--GTSIMPVNPVWAGDVWVPVEGGLFHSIDFGASFTRVGTANATLVS	680
Qy	657	VGFGKS----APGSSYPAVFVVGTT--IGGVTCAYRSDDCGTTWVLINDDDHQYGNWGOAI	710
Db	681	VGAPKSKSDGKKASAPSAVFIMGTDKPGSDIGLYRSDDNGSTWTFVNDQEHYSG-PTWI	739
Qy	711	TGDHANLRVYIGTNGRGIVYDI	734
Db	740	EADPKVYGRVYLTNGRGIVYADL	763

Search completed: May 11, 2004, 12:25:47  
Job time : 90.3468 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 10636.8 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWNSVAIGGGFVD.....YIGTNGRGIVYGDIGGAPSG 740

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgm2\_1/USPTO.spool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORW=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 7742 @runat\_11052004\_114535\_28241 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

RESULT 1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4036	100.0	2869	6	AX700058 Sequence
2	2429.5	60.2	299800	1	AP005028 Streptomy
3	2371	58.7	296500	1	SCO939128 Streptomy
4	2063	51.1	2823	6	AX565635 Sequence
5	2061.5	51.1	3262	1	AF078038 Caldicell
6	2009	49.8	12732	1	AE007608 Clostridi
7	1982	49.1	2950	1	AJ585344 Clostridi
8	1753.5	43.4	2579	8	AB1292929 Agarius
9	1722	42.7	2745	8	AB015511 Aspergill
10	1620	40.1	3959	8	AY040839 Aspergill
11	1579	39.1	2517	8	AY281371 Hypocrea
12	1560.5	38.7	2724	8	AK110506 Oriza sat
13	1405	34.8	299550	1	AP005031 Streptomy
14	1162.5	28.8	14520	1	AE011809 Xanthomon
15	1160.5	28.8	13278	1	AE001712 Thermotog
16	1145	28.4	11910	1	AE012276 Xanthomon
17	1117	27.7	2367	6	AX924431 Sequence
18	1117	27.7	2646	6	AX924429 Sequence
19	1117	27.7	2716	8	AB089343 Geotrichu
20	1114	27.6	2481	6	AX924435 Sequence
21	460.5	11.4	590	8	AB1534359 Agarius
22	368.5	9.1	24208	1	AY281358 Unculture
23	271.5	6.7	47537	7	AY328853 Vibrio pa
24	268.5	6.7	301950	1	AP006570 Gloeobact
25	262	6.5	58128	7	AY357582 Burkholde
26	261	6.5	49575	7	AY328852 Vibrio pa
27	238	5.9	20342	1	AE004613 Pseudomon
28	226.5	5.6	300134	1	AE017211 Geobacter
29	218.5	5.4	7938	1	AE011079 Methanosa
30	214	5.3	5600	1	AF182518 Pseudomon
31	214	5.3	303226	1	AE016774 Pseudomon
32	206	5.1	348866	1	BX640426 Bordetell
33	204.5	5.1	306050	1	RME603645 Rhizobium
34	198.5	4.9	349497	1	BX640440 Bordetell
35	193.5	4.8	348257	1	BX640425 Bordetell
36	193	4.8	303400	1	AP006578 Gloeobact
37	192.5	4.8	207212	2	AL671876 Mus muscu
38	192	4.8	346362	1	BX640439 Bordetell
39	192	4.8	342416	1	BX842573 Mycobacte
40	189	4.7	150894	1	D90904 Synchocyst
41	186.5	4.6	130001	1	D90915 Synchocyst
42	186	4.6	212199	9	AL445199 Human DNA
43	186	4.6	306650	1	AP005221 Corynebac
44	185.5	4.6	2454	4	AF178428 Bos tauru
45	185.5	4.6	4161	1	BSF133614 Bacillus

ALIGNMENTS

AX700058  
 LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003  
 DEFINITION Sequence 2 from Patent WO03012090.  
 ACCESSION AX700058  
 VERSION AX700058.1 GI:29536021  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Acidothermus cellulolyticus  
 Acidothermus cellulolyticus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Frankineae; Acidothermaceae; Acidothermus.  
 1  
 REFERENCE  
 1 Ding, S. Y., Adney, W. S., Vinzant, T. B. and Himmel, M. E.  
 Thermal tolerant avicelase from Acidothermus cellulolyticus  
 Patent: WO 03012090-A 2 13-FEB-2003;  
 Midwest Research Institute (US)  
 FEATURES  
 Location/Qualifiers  
 1..2869  
 /organism="Acidothermus cellulolyticus"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:28049"  
 ORIGIN  
 Alignment Scores: 2.48e-198 Length: 2869  
 Pred. No.: 4036.00 Matches: 740  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 100.00% Gaps: 0  
 DB: 6  
 US-09-917-376-3 (1-740) x AX700058 (1-2869)  
 Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
 Db 139 GCGACGACTCAGCGGTACCTCGGAGCAACGTGGGATCGGGGGCGGCTTGTGAC 198  
 Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40  
 Db 199 GGGATCGTCTCAATGAAGGTGCACCGGGAATCTGTACGTGGGACGACATCGGGGG 258  
 Qy 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db 259 ATGTATCGATGGGATGCGGCAACGGCGGTGGATCCCTCTTCTGGATGGGTGGATGG 318  
 Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 Db 319 AACAAATGGGGGTACACGGGTCTGTACGATTCGGGACACCGGATCAATACTACAAAG 378  
 Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 Db 379 GTATGGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCCAACAGACGGAGCGATTCTC 438  
 Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 Db 439 CGTCTGTCTGATCAGGCGCAACGTGGCAATAACGCCCTCGCTTCAAGCTTGGGGCGC 498  
 Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 Db 499 AACATGCCCGGGCGTGGAAATGGGAGCGGCTTCGGTGGATCCAAACAATGACAAACATT 558  
 Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
 Db 559 CTGTAATTCGGCGCCCGGAGCGCAAGGGCTCTGGAGAGACACAGATTCCGGCGCGACC 618  
 Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAlaProThrAspThr 180  
 Db 619 TGGTCCACAGATGCAAACTTCCGGAAGTGGGACGCTACATTGCCAAATCCCACTGACACG 678  
 Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200  
 Db 679 ACCGGCTATCAGACGCGATATTCAGGGCGTCTTGGGTCTGCTTCGACAAAGTCTTCGTCA 738  
 Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220

Db 739 TCGCTCGGGCAAGCGAGTAAAGCAACATTTTGTGGCGGTGGCGGATCCCAATTAATCCGGTC 798  
 Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
 Db 799 TTCTGGAGACAGACGGCGCGCGAGCTGGCAGCGGTGCCGGGTGGCGGACCGGCTTC 858  
 Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
 Db 859 ATCCCGCACAGGGCGTCTTTGACCCGCTAACACAGCTGCTCTATATTGCCACCAAGCAAT 918  
 Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
 Db 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGCGACGCTCGGAAATCTCGGTGACCTCCGGG 978  
 Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
 Db 979 ACATGGACGCGAATCAGCCCGGTACTCTGACGGACACGGCCACGACTACTTTGGTTAC 1038  
 Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 Db 1039 AGCGGCTCACTATCGACCGCCAGCACCGAATAATGGTGGCAACCCAGATATCG 1098  
 Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
 Db 1099 TGGTGGCGGACACCATTAATCTTTCGGAGCACCGCGCGGTGGCAGCTGGACGGGATC 1158  
 Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
 Db 1159 TGGGATTTGGACGAGTATCCCAATCGAAGCTTGCATATGTGTGACATTTCCGGGAG 1218  
 Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
 Db 1219 CTTTGGCTGACCTTTCGGGTACAGCCGAATCTCCCGTACCAGTCCGAAGCTCGGCTGG 1278  
 Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 Db 1279 ATGGATGAAGCGATGGCAATCGATCCGTCAACTCTGTATCGGATCTCTACGGAAACAGC 1338  
 Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 Db 1339 GCGAGCTTTGTACGCAACAATGATCTCAGAGTGGGACTCCGGCGGCGCAGATTCATATC 1398  
 Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440  
 Db 1399 GCGCCGATGGTCAAAAGGATTCGAGGAGACGGCGGTAAACGATCTCATCAGCCCGCGTCT 1458  
 Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
 Db 1459 GGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGGCGGCTTCCACCACCGCCGCTTACT 1518  
 Qy 461 AlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAspTyr 480  
 Db 1519 GCCGTGCATCGAGATCTTCACGTACCGGTGTTTCAGACCGGCGCAGCGCTGCACTAT 1578  
 Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
 Db 1579 GCGGAATTGAATCGTCCGATCATCGTTCGCGCTGGAAGTTTCGATCCATCCAGCAACCG 1638  
 Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGlu 520  
 Db 1639 AACGACAGGACGCTCGCGTCTTCGACACAGCGCGCAAGAACTGGTTCACAGCGACGCA 1698  
 Qy 521 ProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
 Db 1699 CCTGGCGGGGTGACGACGGCGGACCGCTCGCGCATCGCGCGACGGCTCTCGTTTCGTC 1758  
 Qy 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560  
 Db 1759 TGGGCTCCCGCGGATCCCGGTACGCTGTGTGTACGAGTCCGGATTTGGCAACTCTCTGG 1818  
 Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580

Db 1819 GCTGCTTCCAGAGGTGTTCCCGCCCAATGCCCAATCCGCTCAGACCGGGTGAATCCAAAG 1878  
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
Db 1879 ACTTTCTATGCTCCTATCCAAATGAACCTTCTATCGAAGCAGCGCGCGGTGACATTC 1938  
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620  
Db 1939 CAACCGGTGCGGCGCGGTCTTCGAGCAGCGGTGCGGTGCTCATGTTCCACGCGGTG 1998  
QY 621 ProGlyLeuGlyGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640  
Db 1999 CTGGAAGAAAGAGCGCATCTGTGCTCGCTCGCTCATCGAGCGGGCTTTACCACTCAACCAAT 2058  
QY 641 GlyCysSerSerTrpSerAlaThrGlyValSerSerAlaValAlaValGlyPheGly 660  
Db 2059 GCGCGCAGCAGTTGGTCTCAATCACCGCGGTATCTCCGCGGTGAACGCGGATTTGGT 2118  
QY 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrLeuGlyGlyVal 680  
Db 2119 AAGTCTGCGCGCGGTGCTCATACCCAGCGGTCTTTGCTCGCGCAGCATCGAGCGGT 2178  
QY 681 ThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeuLeuAsnAspGln 700  
Db 2179 ACGGGGCGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGATCAATGATGACCAG 2238  
QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaThrGlyAspHisAlaValArgVal 720  
Db 2239 CACCAATACGGAATTTGGGACAAAGCAATCACCGGTGACCAACGCAATTTACGCGCGGTG 2298  
QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 2299 TACATAGGCACGAACGCGCGTGAATGTATACGGGACATTTGGTGTGCGCGTCCGGA 2358

RESULT 2  
AP005028 299800 bp DNA linear BCT 10-MAY-2003  
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 8/30.  
DEFINITION AP005028 BA000030  
VERSION AP005028.1 GI:29605366  
KEYWORDS Streptomyces avermitilis MA-4680  
SOURCE Streptomyces avermitilis MA-4680  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
MEDLINE 21477403  
PUBMED 11572948

REFERENCE 2  
AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.  
TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)  
MEDLINE 22608306  
PUBMED 12692562

REFERENCE 3  
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., and Hattori, M.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0086, Japan

(E-mail: bioelite.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda (\*1), Jun Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi Osonoe (\*4), Norihiro Kishida (\*4), Hisashi Kikuchi (\*4), Tadayoshi Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Masahira Hattori (\*1, \*7) and Satoshi Omura (\*1, \*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermitilis.ls.kitasato-u.ac.jp.  
Location/Qualifiers  
1. 299800  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"  
16. 918  
/note="SAV1722"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC69433.1"  
/db\_xref="GI:29605367"  
/translation="MAGRALAVPCADLTLPTMSAESDGPGRVDRLLRYPVKSMLEE LPAAVEVERGLAHDHRRVALIDRTGKVASAKHPRLMSRLLTVTADNDSAVRMTTSDG KEIWDTPGVDEALSEIVGPEVLTADPPHEALDRADPEVLRDHTRPVPSARTGRI GGSGPGETFPFAPVHWTTSTLARIALAPNGSAQERYRNI VLRTPLTRGFTEND WIDRTHLGTGLVLRVITARTPRCAIPTLAHGTVPADTALRLVLAENHRTVPTESMGPG PCAGVYAQLRPGRIIRLGDVPHWA"  
complement (1232..3406)  
/note="SAV1723"  
/codon\_start=1  
/transl\_table=1  
/product="putative mycodextranase"  
/protein\_id="BAC69434.1"  
/db\_xref="GI:29605368"  
/translation="MHSNLSRYVRMSAIGAAVALAAGLLVALAPAAHAAAGADLPFT SVEASATTTGKTGIPDHTQCTIASEASGROAVRLGAGORVFTVPRANAVNVSVN PDGSGALNVTVNGTKLAKTLAVTSKSYVDTSWIGAKTHFFDNARILLGNVQPG DKIAVEAVSTQVTVNADFEQVAAATKPAVASVTDKADPGSGDSDTOAFNAAIA AQGGVWIPGDRYLTSLNGVNTLQAGSWHSVHTSRFIDSSSGGSHLKDFA VIGBTVRSDNPDNFVNGSLGPGSSVGMWLKVLWLTGNDNLVNNRILDT TADGLNNGARGVRNRNLRNGDDALAMWSLYSPDNTSSFSNTISQPNLANGIA IYGGTDLAKRNLSIDTALSGSLAIKQKELDPESPLAGTITVSGNTLVRAGMNP WQHPMGALRSDSYDAEANSVITGTTIIDSFYSAFEFVSGGCHGTATKNTVTDGATV RNTGTVVVOASQGAKEFNVQATGVGAAGIYCPYPAGSGFTTLDGNGNAGWSST GDCSAMPQCGQNPDPPTRNLAKRPAATGSDQVYTPGKAGDADSTYWGSTNNAF POAWTVDLGSSQAVRLVRLKLPATAMQARTQTLVSGSTDCGSSTVVASQGYRDFP ATGNTATVTPSGTNLRVLRKLVNTANTCWPAAQFSEVEAYLS"  
complement (3464..5065)  
/gene="aglA1"  
complement (3464..5065)  
/note="aglA1"  
/gene="SAV1724"  
/note="SAV1724"  
/codon\_start=1  
/transl\_table=1  
/product="putative alpha-glucosidase"  
/protein\_id="BAC69435.1"  
/db\_xref="GI:29605369"  
/translation="MGQPTARIQKDWRTAVIYQVYVRSFADGSDGTGDLAGVRAR LPYLAGLVDAIWFPNFWLSPMKDGYDVADYRAVDPAFGTLAEAEKLTAEARELGR





Db	176418	GC	GGGCGGCGACCTGGTCCCGGCTCGCGCGCCAGCCACCGGCCATCTCGCCCAACAAGGCGC	176477
Qy	246	Val	PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr	265
Db	176478	GT	GTGACGCGCGGCGGCGGCTGTCTGTACTCTCGGTACACGCAAGGCGCGACCGTAC	176537
Qy	266	Asp	GlySerSerGlyAspValTyrPheSerValThrSerGlyThrTyrThrArgile	285
Db	176538	GAC	GGCGCAAGGGGACAGCTGTGGCGGTACACGACGAAGACCGGGACCTGGACGAACATC	176597
Qy	286	Ser	ProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrile	305
Db	176598	AGC	CGGTCGCGGAGCGGCGACAC-----TACTACGGCTTCAGCGGCTGACCGTG	176648
Qy	306	Asp	ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTyrProAspThr	325
Db	176649	GAC	CGGCGAGCATCCGGGCGAGCGTATGGCGACTCGGTACAGCTCCTGGTGGCGGACAGC	176708
Qy	326	Ile	IlePheArgSerThrAspGlyGlyAlaThrTyrThrArgIleTyrAspTyrThrSer	345
Db	176709	CAG	CTCTTCGCTCCACGGACAGCGGCGGCACTTGGACGAGGCGCTGGGACTACACCTCG	176768
Qy	346	Tyr	ProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrTyrLeuThrPhe	365
Db	176769	TAT	CCGAGCGCGCTCGAAACCGCTTCCACATGGATGCTCTGCTCGCTCGCTCGCTCGCTCG	176828
Qy	366	Gly	ValGlnProAsnProValProSerProIysLeuGlyTyrMetAspGluAlaMet	385
Db	176829	GG	ACGAAACCCCGCACCGCGCGACGACACCCGAACTCGGCTGGATGACCGAGTCCCTG	176888
Qy	386	Ala	IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405
Db	176889	GAG	ATCGACCGCTTCGACTCGCGCGCATGATGTACGGAACGGCGGCGGCGCTCTACGGC	176948
Qy	406	Thr	AsnAspLeuThrIlystTyrAspSerGlyGlyGlnIleHisIleAlaProMetValIys	425
Db	176949	ACG	GACAACTGACAACTGGGACAGCGGAAGCCAGTTCACCATCAAGCCGATGGCGCGG	177008
Qy	426	Gly	LeuGluGlnThrAlaValAsnAspIleSerProProSer---GlyAlaProLeu	444
Db	177009	GGC	CTGAGAGAGACGGCGCGTCAACGACCTCGCTCGCTCGCGCGCGCGCGCGCGCTG	177068
Qy	445	Ile	SerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Db	177069	TT	ACGCGGCTCGGTGACATCGCGCGGCTTCGGGCACACGACCTCACCGTGGCGCTCG	177128
Qy	465	Thr	IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAla	484
Db	177129	CT	GATGTACAGCTCGCGCACTTCACCACGAGCACCGCTCGACTACGCCGAGACCGAC	177188
Qy	485	Pro	SerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Db	177189	CCG	GGCAGGTGGTGGCGGTGGGCAATCTCGAC-----TCGGGTCCG-----CAT	177233
Qy	505	Val	AlaPheSerThrAspGlyGlyLysAsnTyrPheGlnGlySerGluProGlyGlyVal	524
Db	177234	GT	GCGTTCGACGACACCGCGCCCACTGGTTTCGGGGGGCGGACCTTCGGGGGTC	177293
Qy	525	Thr	ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly	544
Db	177294	AGC	GGGGTGGGACCGTCCGCGCGCGTCCGACGCGAGTGGCTCGTGTGGAGCCGCGCG	177353
Qy	545	Asp	ProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTyrTyrAlaAlaSerGln	564
Db	177354	GGC	ACCGGG-----GTGCGAGTACACACCGGGTTCGGCACCTCGTGGTGGCGTCCGCG	177407
Qy	565	Gly	ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Db	177408	GGC	CTCCGCGGGGGCGATCGTCACTCCGACCGGGTCGACCCGAGACCTTCTACGGC	177467
Qy	585	Leu	SerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla	604

Db	177468	TTCAAGTCCCGGACAGTTCTTACGTTCAGTTTCGACGCGGGGGCGACCTTCACGCGCTGCCGCG	177522
Qy	605	Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLylys	623
Db	177528	GCCACGGGCGCTGCCGAGCGGCGACAGC-----GTGGCGCTTCAAGGCGCTGCCGCGCAGC	177581
Qy	624	GlulGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Db	177582	AAGGGCGACATCTGGCTGGCGGCGCGAGCGACGCGCGGTACGGGCTGTGGCACTCG	177641
Qy	639	ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Db	177642	ACGACGCGCGCGCGGCGCTTCCACCAAGCTCCGCCCGTCGACAGCGCTCTACACGAGCGGAGATCGGC	177701
Qy	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVaIcIleThrIleGly	678
Db	177702	TTGCGAAGGCGCGACCGCGCGCTCTGTACAGACGCTCTACACGAGCGGAGATCGGC	177761
Qy	679	GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp	698
Db	177762	GGTGTGCGGCGCATCTCCGCTCGACCGCAACAGGCGCGAGCTGGACCCGCGTCAACGAC	177821
Qy	699	AspGlnHisGlnTyrGlyAsnTrpGlyClnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	177822	GATCCCAACCACTGGGGTTGACGCGGCGCGGATCAACCGGTGACCCCGAGGCTTCACGGG	177881
Qy	719	ArgValTyrIleGlyThrAsnGlyVArgGlyIleValTyrGlyAspIleGlyGlyAlaPro	738
Db	177882	CGCGTGATGTGTCTCGACGACGCGGCGGGATGCTTACGCGGACACCGCGCGCTTCCTCG	177941
Qy	739	SerGly 740	
Db	177942	GACGCG 177947	
RESULT 3	SCO939128/c		
LOCUS	SCO939128	296500 bp	DNA linear BCT 11-FEB-2003
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 25/29.		
ACCESSION	AL939128	AL023862	AL031013
VERSION	AL033505	AL034443	AL035161
KEYWORDS	AL939128.1	GI:24413907	
SOURCE	Streptomyces coelicolor A3(2)		
ORGANISM	Streptomyces coelicolor A3(2)		
REFERENCE	Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;		
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.		
1	Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,		
	Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,		
	Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,		
	Crozin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,		
	Huang, C.H., Kieser, T., Lark, J., Murphy, L., Oliver, K., O'Neil, S.,		
	Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,		
	Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,		
	Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrell, B.G.,		
	Parkhill, J. and Hopwood, D.A.		
TITLE	Complete genome sequence of the model actinomycete Streptomyces		
JOURNAL	coelicolor A3(2)		
MEDLINE	Nature 417 (6885), 141-147 (2002)		
PUBMED	21996410		
REFERENCE	2 (bases 1 to 296500)		
AUTHORS	Bentley, S.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces		
	sequencing team, Sanger Institute, Wellcome Trust Genome Campus,		
	Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk		
COMMENT	On or before Oct 26, 2002 this sequence version replaced		
	gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757,		
	gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765,		
	gi:20520767.		
FEATURES	Location/Qualifiers		
source	1. 296500		

```

/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
complement(144..443)
/gene="SCO6480"
/notes="synonym: SC9C7.16c"
complement(144..443)
/gene="SCO6480"
/notes="SC9C7.16c, unknown, len: 99 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.16c"
/protein_id="CAA22728.1"
/db_xref="GI:4154075"
/db_xref="SPTREMBL:Q9ZBJ4"
/transl_table="MTYGCATSKMTRPGFEEVUVLLLSAADGLRDAGCELYVGLSE
DDSTIHWIVRWTRTDESLDPAKDAIGRAMPLTGEFTKQMSVAGGLGL"
complement(517..1563)
/gene="SCO6481"
/notes="synonym: SC9C7.17c"
complement(517..1563)
/gene="SCO6481"
/notes="SC9C7.17c, conserved hypothetical protein, len:
348aa; similar to several hypothetical proteins eg.
SM:Y0BN MYCTU from Mycobacterium tuberculosis (340 aa)
fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.6%
identity in 355 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22729.1"
/db_xref="GI:4154076"
/db_xref="SPTREMBL:Q9ZBJ3"
/transl_table="MRWENLTDSDHARAADTALFGADVTRFDTEPRGIFTHEVR
ARSLNRVPGASRMPEFWTVNYRGCTHACVYCFARKTHSYLDDTGLGFDSDQIVVKV
NADLEVRQRQASRWQEHVAMTNDVCYQRAEGRVLMFGIIIEALTRANPFSILTK
GTLIRDLDLTLRAARVTEGIVSVSGFTDAELWRTVEPTGAPARLEVRVLRGEG
IGCVLMAPVPIPLFGDEPAOLRATVRAIAAAGATVTPVLHLRPGAREFWMAWLGRR
HPLHVRERYLYADGAVAPKWYQRRITRQVHDLAREVIGIGPARAGMRRVAEPHSRP
AEFAPPEPVQLSLI"
1800..2243
/gene="SCO6482"
/notes="synonym: SC9C7.18"
1800..2243
/gene="SCO6482"
/notes="SC9C7.18, conserved hypothetical protein, len: 348
aa; similar to two hypothetical proteins e.g. TR:069685
(EMBL:AL022121) Mycobacterium tuberculosis RV03718c or
MTW025.066c, 147 aa; fasta scores; opt: 475, z-score:
768.3, E(): 0, (52.1% identity in 146 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22730.1"
/db_xref="GI:4154077"
/db_xref="SPTREMBL:Q9ZBJ2"
/transl_table="MAQVEATYTERIVADAETVFDLADYSGTREKLLPHEPSEYVR
EGDGGTGLVHWKLTQSKRVRCDCLEVEFTDGLVEKDRNMSMTVTRVTPAGEGK
SRVVLVTWQAGVGCGGFFKTFAPKGLARIYDALLAKLAAEVEK"
2394..4841
/gene="SCO6483"
/notes="synonym: SC9C7.19"
2394..4841
/gene="SCO6483"
/notes="SC9C7.19, probable efflux protein, len: 815aa;
similar to many, several of which are involved in toxin
antibiotic efflux eg. TR:069070 (EMBL:AF061335)
oxytetracycline exporter from Streptomyces rimosus (536
aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,
(36.9% identity in 542 aa overlap)"
/codon_start=1
/transl_table=11

```

```

/product="putative efflux protein"
/protein_id="CAA22731.1"
/db_xref="GI:4154078"
/db_xref="GOA:Q9ZBJ1"
/db_xref="SPTREMBL:Q9ZBJ1"
/transl_table="MGGITLVQDPVRDPQDEPVVAAAPPGPPVPGDGERDAGLSP
RRVLGVFLGRLMALLALEQMIIVATAPKIVGELHGLDRMSWALTALDILSTVGUPI
YKGLDLFGRLKGVQFAILVFVVGSLAGRAGTMDQIAFRAVAGGAGLGMIGVOAI
IADIVPRERARYMGLIGASGLASVAGPLGSGYFTDLSWRWCYFINVFGLVTVAV
VAVVLLKLPVRPRKLDIIGALLIAAASCTCLVLTSMGCTEYAWGSRVILGAAAVA
ATLLFLVAEHPAPEPLIPLRFRDSIENVATLVGLVGVALFGAASLYLPTFLQMDVGA
SATESGLMLPMGCVGASIIISQLISRTGCHYRIHPILGALSIVGVWLLSRLDADT
SRHLSIWMVAMVIGAGLIGLVMPVLVAVQNSVRITDGTATGATANNVFRQIGSGRAVF
GTLFAGRLTADALADRIPAEAGVGLPDAEAITPQLVHSLPPALRDYIIGAYDAMPRIIF
LYLVPLVLIGLLIALIFLKEPLVSHNAPVTPETAHPAQPPQAPQVPPARSPYAGGI
PVCTVOHPDGTGVVPRAALTILIDVGQOIGRSGDGDGRYALATPGSGAYVLIAAAGG
HQPQAVSVTVGGERPVELDVILGAGRLAGSVMTAGDSPVRDATVTLTNVHGVRVATTR
SGRGGGVITELVAGEYITLAAAFAPRAALPVSVQAAARETRQDVVELAGGAVLRTVR
ASGGRAVEDARVTLDDAAGNVVDLTITGPDGTFRFRVLDSSGEYTVIAAGYPVATVQLQ
VAGGRTERDILQGHED"
4911..7493
/gene="SCO6484"
/notes="synonym: SC9C7.20"
4911..7493
/gene="SCO6484"
/notes="SC9C7.20, conserved hypothetical protein, len:
860aa; similar to many hypothetical proteins eg.
TR:086875 (EMBL:U22894) from Streptomyces lividans (547
aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,
(40.0% identity in 497 aa overlap) and TR:086807
(EMBL:AL031031) from Streptomyces coelicolor (916 aa)
fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
(27.2% identity in 779 aa overlap). Also contains a short
region of similarity to TR:Q53897 (EMBL:X60316) AbaA
regulatory locus from Streptomyces coelicolor (192 aa)
fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
(38.5% identity in 135 aa overlap)".
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22732.1"
/db_xref="GI:4154079"
/db_xref="GOA:Q9ZBJ0"
/db_xref="SPTREMBL:Q9ZBJ0"
/transl_table="MVTGHPRTVTVATQILRSRQGRGPDAMDHGTGTRDAPPGRAGPD
TVPADPVTGRVPLAVVDRDGLYSHMSRGARLFLALKEAIGEPAPDLLPVSGLVP
EDADPDGAGHDDSTRSTADDPGPGFLESSLDGRISYPAGARLITVRADRDVLMW
AYPLVPGPERLLVLAADADGLRDYADGTDSDAGPAVERIAPAFALHTDPPGDEL
ARRLPEILPMSVGESARIVAQVLELGYVPLEFSQHDRVPTVDMGVPRRAERARE
RAARLAAAGEPVPDELDEADLEAVRERLEFLNEVSGAIGTSLDLSRTIVEVSRA
VVPRTDVAGTYLREQVAVGPGDGVDDTTTMMHVALEHTDEPRDWDVVVGEAM
PPAHTTFFQCMTTEGPVLPRIETEMGHAIASQFEKDIIRPLTGRSMILVPLKART
VVLGFMILLRHPERPVPNDMDRVTGAEALAAAGLVLDNARMYTHQENVATLQDSMLP
HIPARLAGCDIATRYLPTLGRVGGDFDAVLPKLPARTLVVGVDMGHGLNSAAMWG
QLRTAVQTMAGLDLPPAQLLNLDLQALRGDTHTLATCLYAYDPIAGELHLANAGHI
PPLVRAEDGASLIDLTPTAGIPGVGVFPFESVTVVAPDRLVNCTDGLVEVRGEDI
GVGLATICESAAHPAASMDACDTIIRALGAGGGRKDDVALLMARLTGTIEPDVAEWR
LARDPVRGRARAAREQLYDWGLPKLVAPRELMSVELTVNAVRHSHARPVALRLVRA
DTLLCEVDVDDDDHLPALKSAGPDEGTGRLRVSVTLAREWGASRTGAGKTVWFELTP
ARRH"
7711..8433
/gene="SCO6485"
/notes="synonym: SC9C7.21"
7711..8433
/gene="SCO6485"
/notes="SC9C7.21, unknown, len: 240 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.21"
/protein_id="CAA22733.1"
/db_xref="GI:4154080"
/db_xref="GOA:Q9ZBI9"
/db_xref="SPTREMBL:Q9ZBI9"

```

gene  
 /translation="MSVTSRYRDWEGFWNEASGDRGAVFWDAEPTLTAGRHILAHFPP  
 HLTDALPLDLGCGNGTQRYFAERPHVVGADLSAALAHARHADPPGLAAYRQLD  
 AADKAQALHAGLGDANVTYRVGLHQCEPDQROFLADALAALIGRGRFLVPEPSQA  
 AKTVLMGLAQCPAGPPAKLAPLVRHLGLTPGAVPDEAVPEYLRAAGLTVLASEGELPLVT  
 TKTADGTRLELPSAWVAGTA"  
 8603. 9436  
 /genes="SC06486"  
 /note="synonym: SC9C7.22"  
 8603. 9436  
 /genes="SC06486"  
 /note="SC9C7.22, possible transport associated protein,  
 len: 277aa; similar to SW:DPFA\_BACSU dppA, dipeptide  
 transport protein (274 aa) fasta scores; opt: 448,  
 z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa  
 overlap)."  
 /codon\_start=1  
 /transl\_table=11

## Alignment Scores:

Pred. No.:	1.24e-110	Length:	296500
Score:	2371.00	Matches:	431
Percent Similarity:	72.29%	Conservative:	109
Best Local Similarity:	57.70%	Mismatches:	183
Query Match:	58.75%	Indels:	24
DB:	1	Gaps:	10

US-09-917-376-3 (1-740) x SCO939128 (1-296500)

Qy	6	TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn	25
Db	70439	TACACCTGGGAAGACGCCGATCGAGCGGCGGCTTCGTCGCCGCGATCGTCTTCAAC	70380
Qy	26	GlucylalaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	45
Db	70379	CGGACGAGGAAGACCTGGCGCTAGCCCGCACCCACATCGGAGCGGCTACCGCTGGGAG	70320
Qy	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65
Db	70319	GAGAGTCCACACCTGGAGCGCGCTCTCGACACATCGTGGCTGGAGACTGGGGGAC	70260
Qy	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaAlaVal	85
Db	70259	ACGGCGTGTGCTCGCGCTCGACCGCGCTCGATCGGACCGGCTACCGCGCGGCTC	70200
Qy	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105
Db	70199	GGCAGTACACCAACGACTGGGACCCGACCAACGCGCGGCTGCGCTCGCGACCGG	70140
Qy	106	GlyAlaThrTrpGlnIleThrProLeuProPheIleValLeuGlyGlyAsnMetProGlyArg	125
Db	70139	GGTGGAGCTGGAGAGGCGGACCTCGGCTTCAAGCTGGGCGGCAACATGCCGCGCGG	70080
Qy	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	70079	GGCATGGTGGAGCGCTGGCGCTCGACCGCACCAACGACGCTGCTACCTGGCGGCC	70020
Qy	146	ProSerGlyIleGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	165
Db	70019	CCGACGGCCACGACTGTGGCGGTCCACGACGCGGCGGCTCACCTGGTCGAGGTGACG	69960
Qy	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	185
Db	69959	GCCTTCCGAAACCCGGGAACCTAGCCGAGACCCGACCAACGACGTCGGCTACGCTCC	69900
Qy	186	AspIleGlnGlyValValTrpValAlaPheAspIleSerSerSerSerLeuGlyGln	204
Db	69899	GACAAACGAGGCATCACTGGGTCACTTCGACGAGTCGACGCGGCGGCGGCGGACG	69840
Qy	205	AlaSerIleThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg	224
Db	69839	GCCACGGGACCTCTACGTGGGGTTCGCGACAAAGGAGAACGCGCTCTACCGCTCGACC	69780
Qy	225	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisIys	244

Db	69779	GACCGGCGCGACCTGGGAGCGGCTCGCGGCGACCGCCGCGGTTACCTGGGCCCAAG	69720
Qy	245	GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro	264
Db	69719	GGTGTGCTGGAGCGGAGAACCGCTACCTGTACTCGCTTACAGCAGCACCGCGGCGCCG	69660
Qy	265	TyrAspGlySerSerGlyAspValTrpIlePheSerValThrSerGlyThrTrpThrArg	284
Db	69659	TACGACGCGGCAAGCGCGGCTGTACCGGTACGCGACGCGGACCGGACCTGGAGCGAC	69600
Qy	285	IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr	304
Db	69599	ATCAGCCCGCGCGGAGCGGACACCTACCTACGCGCTTACGCGCTTACGCGCTGACG	69549
Qy	305	IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerThrTrpProAsp	324
Db	69548	GTCCGACCGGAGCGGCGGAGCGGTGAGGCGACCGCGTACAGCTCGTGGTGGCGGAG	69489
Qy	325	ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr	344
Db	69488	ACGCAGATCTTCGCTCGACGCGACGCGGCGGACATGTCCAGCGGTGGAGCTACACC	69429
Qy	345	SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr	364
Db	69428	TCGTACCGGACCGCGAGAACCGCTACAGATGAGCGCTCTCTCGTCCCTCGCTCACC	69369
Qy	365	PheGlyValGlnProAsnProValProSerProIleLeuGlyTrpMetAspGluAla	384
Db	69368	TGGGCGCGAACCCTCGCGCGCGCGGACGACACCCCGAAGCTGGGCTGGATGACGAGGCT	69309
Qy	385	MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr	404
Db	69308	CTGGAGATCGACCGCTTGCATTCGACCGGAGTATGTACGCGCGGCGGCGGCTAC	69249
Qy	405	AlaThrAsnAspLeuThrIleTrp---AspSerGlyGlyGlnIleHisIleAlaProMet	423
Db	69248	GGCAGGAGAACCTGACGAACTGGGACGACGAGGCGGCGACGTTTCGCGCTCGGCGGATG	69189
Qy	424	VallysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaPro	443
Db	69188	GTGCGGGCGCTGGAGGAGACGCGCTCAACGACCTGGCGCTCCCGCGCTCGGCTGCGCGG	69129
Qy	444	LeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValPro	463
Db	69128	CTGCTCAGCGCGCTGGTGACGTCGCGGCTTCGCGGCGACACGACCTGACCGAGGTCGCG	69069
Qy	464	SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu	483
Db	69068	TCGATGATGTACACCTCGCGGAACTTCACCTCGACGACGAGCGCTGGACTTCGCGGAGACG	69009
Qy	484	AsnProSerIleValArgAlaGlySerPheAspProSerSerSerGlnProAsnAspArg	503
Db	69008	AAGCGGACGCTGTGTCGCGGCGGCAATCTCGAC-----TCCGCTCG-----	68964
Qy	504	HisValAlaPheSerThrAspGlyGlyIleAsnTrpPheGlnGlySerGluProGlyGly	523
Db	68963	CACATCGCTTCTCCACGACCAACGCGCGCAACTGTTTGGCGGCGACCGGCTCGCGG	68904
Qy	524	ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro	543
Db	68903	GTGAGCGGCGGCGACGCTCGCGGCGGCGGAGCGGCGGCGGCTTCGTTGTCGAGCGCG	68844
Qy	544	GlyAspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaIleSer	563
Db	68843	-----GAGGCTCGCGGCTGAGTACACGACCGGCTTCGCTCGGCGGCGGCTCG	68790
Qy	564	GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProIleThrPheTyr	583
Db	68789	ACCGCGCTCCCGGCGGCGGATCGTGGAGTCGAGCGGCGGCTGAAACCGCGGACCTTCTAC	68730
Qy	584	AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal	603



```
Db 901 TATGACGGGTCGAAGGGGATGTGGCGCTAGATATTTCAGTGGCCAGTGACTCGA 960
Qy 285 IleserProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304
Db 961 ATCAGCCCAATTCCTCTACC---TCATCGAATTCACGGTTCGGGTACTCCGGTTAGCC 1017
Qy 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTrpProAsp 324
Db 1018 ATTGACCGCAAAACCTGACACCATCATGGTTGTTTCCCAAGTGTCTTGGTGGCTGAC 1077
Qy 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTyrAspTrpThr 344
Db 1078 ATGTACGTATATGCTCCACCGACCGGGGAAAACGTGGTGCCTATCTGGGAACCTGAAC 1137
Qy 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364
Db 1138 GGTTCCAGCGCGGTACAAAGCAGTACACACCATTTACTCAGGGGTCTCCCTGGCTTGAC 1197
Qy 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTyrMetAspGluAla 384
Db 1198 TTGGCAATACAGCAAGAGCCTGAAGCTAACCCCAAACTCGGGTGGATGACCCAGTCC 1257
Qy 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
Db 1258 TTCGAGATTGACCGCATACTCCGACCGCTTCTTCTATGGCACCGCGGGCATCTAT 1317
Qy 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetVal 424
Db 1318 GCGCGAAGCAACCTTCACAACTGGGTAAGGGCAAAAGTTGATATCATTCTCAAGGCC 1377
Qy 425 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444
Db 1378 CAAGGGATCGAAGAAACCGGACGCCAAGACTTCGCCGCTCCCTCCCTGGAAACATTGGCCCTA 1437
Qy 445 IleserAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464
Db 1438 TACTCAGCGTGGCAGACATCGTGGTCTCACCACAAAGATATCTCAGTGCCTAAC 1497
Qy 465 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483
Db 1498 AAGTACTACTACAAAGACCCACACCATGACACCGTGCCTCCATTGACTTCGCCGAAAGC 1557
Qy 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
Db 1558 AAACCGGCAACTGCTGCGTGGCGAGGCAAA-----TCTATTCTTGGGAAACACCACTCA 1611
Qy 504 HisValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGluProGlyGly 523
Db 1612 TGGGTGGGGTTTCACTGATGCTGGCGAAGCTGGAAACCTGGCGCCACGCCATCAGGG 1671
Qy 524 ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
Db 1672 GTTAAAGCCCGCGCTCCATCCTGCTGCTAGCTAATGCTTCATCCTGTTGTGGGCCCG 1731
Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
Db 1732 -----GAGGTGGCGCTCTCTCCAGCTTCCACCAACAGTGGCTCTCTCGTGGTCAACGGTG 1785
Qy 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583
Db 1786 AGCGGGTCTCTCAATGCCAGCTGCATCAGCAGAGTCAACGCTAATATCTCTTTAC 1845
Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
Db 1846 GGCTTCGTTGACGGAAGTTCTATCACTCCACAAATGTTGGGGTCTCTCTCCACGCGTCA 1905
Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
Db 1906 GCTTTCACGGGGTTTCCACACCAAGTGA-----AACGTTCTGTTTCCGTCGCGTGCAGGC 1959
Qy 623 LysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637
```

```
Db 1960 GGTCAAGGTCACTCTCGCTGCTGCGAGTATCGGGAAGCACCTATCGGATGTGGCGG 2019
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
Db 2020 TCCACTGATGCGGGAAGACTGGACCAAGGTATCTGGAGTTCCAGGAAGCGGATCGGTT 2079
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677
Db 2080 GGTTCGGTAAAGCAACCAAGTAGTAGCGATACCCAGTGTATTTTCACATCCGCGCAAGATT 2139
Qy 678 GlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsn 697
Db 2140 CATGAGTTCAGGAATTTTCGGTCTGACATGAAGCAAGCAACCGTGGAAACGCATCAAC 2199
Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
Db 2200 GACAACCAACCACTGTTTGGACTGGCGCATCATCACCCTGACCCAGATGTGAA 2259
Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAla 737
Db 2260 GGGCGGCTACATTGGTACCAACCGTCCGCGCATTTATCGTTGGTATTCTCTCCACTCCC 2319
Qy 738 Pro 738
Db 2320 CCG 2322
RESULT 5
AF078038S1
LOCUS Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial cds.
DEFINITION AF078038.1 GI:12743874
ACCESSION AF078038
VERSION AF078038.1
KEYWORDS 1 of 5
SEGMENT
SOURCE Caldicellulosiruptor sp. Tok7B.1
ORGANISM Caldicellulosiruptor sp. Tok7B.1
REFERENCE 1 (bases 1 to 3262)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 20171169
PUBMED 10706665
REFERENCE 2 (bases 1 to 3262)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
FEATURES
source
1. 3262
/organism="Caldicellulosiruptor sp. Tok7B.1"
/mol_type="genomic DNA"
/isolate="Tok7B.1"
/db_xref="taxon:80339"
274. .>3262
/note="ORF5; multidomain"
/codon_start=1
/transl_table=11
/product="glycosyl hydrolase 5"
/protein_id="AAK06388.1"
/db_xref="GI:12743879"
EGGGFTIGKIVFNPKENLYVTRDIGNARSTDSGNTWTQLMWVSFDNNLLGVES
IATDPVDPNKVILACQGTITNSWTMDNGAILRSTDEGDTFEITPLDFXLGGNMPXRL
GERLADIDNNRILYLRBEGKLVSESDYGVSWKKVTSFPNPGTYIEDPNCNDYLN
HITGVVVVFPDTSRPGEGKKIYVGVADKTSIYTKDGGTQWALFQPTGLLPQ
RAKLSDDGLMXYTNSYQPYNGDYGEVRYNTKGWKNISPMAAQDTFYGGGLAV
DAQNPKVVMVAALSSWEDTYIWRSIDGSGTWKCIWENGYENRRTLHYNWDISAAPLV
```

NFGXTBPTPEVSPKLGWVGTLEIDPFNSDXMLYGTGAXLYGCDLITNDKQNTI	
KVAIGIEBTSVQALISPPVPHLFSALGDIAGRHEDLEKAPNWTYVQNMGTDDI	
DFAELNPNMVRVGVNDYDQWNPNTNIRFGSDGKSWFQGNTEPQGTSEGTVA	
GSAVWAPKAGKVCYSDNGKWKVECANVPEAIYVSDRPNPKFYAFKNGKFI	
KGTFTESPAAGLPIISGNPKTVPIEGDIWLGNNGMHSWDGYSFVKISGVEDA	
IGFKPABEYTPAIYIYAKINGVRGIFRSDDCDKTWRINDDKHPCGANADITGP	
RVGRVFNATGILGWGETAYISNISPATSTPSTPSTPPTPTATITSTPATST	
PTPTPTPTATPT	
SAGSIRPFKIVNGSSVDLSVKIRYWTYVDGDKPQSAVCDWAQIGASNVTFN	
LSGVSADYLLVSGFSSGAGOLQPGKDTGDIQVRFNKNDWSNYNQADWSLQ	
YGNNAKVLVVDGVVLWGQEPGGA	
misc_feature	274..2580
/note="Region: endoglucanase domain"	
/evidence=experimental	
misc_feature	2791..>3262
/note="type IIb cellulose binding site"	
misc_feature	2581..2790
/note="PT-linker"	
ORIGIN	
Alignment Scores:	
Pred. No.:	5,85e-97
Score:	2061.50
Percent Similarity:	65.62%
Best Local Similarity:	52.58%
Query Match:	51.08%
DB:	1
US-09-917-376-3 (1-740) x AF078038S1 (1-3262)	
Qy	4 GlnProThrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyLeu 22
Db	376 CAGCCATATGTGTGAAGAATGTTAAATTTGAAGTGGAGGAGGTTTATCATCGGAATA 435
Qy	23 ValPheAsnGluGlyAlaProGlyIleLeuThrValArgThrAspIleGlyMetTyr 42
Db	436 GTGTTTAAACCCGGAAGAGAAAATCTTTGTTTACGTTAGGACAGATATAGGAGGAGCTTAC 495
Qy	43 ArgTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTTrpValGlyTrpAsnAsn 62
Db	496 CGTAGTACAGCGGTGCAATACATGACACAACTTATGGAAGTGGGTTAGTTTGTATGAT 555
Qy	63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal--- 81
Db	556 TGGAAATTTGCTGGGTGAGAGAGCATTTGACAGATCTCTGTGATCTCAATCGGTTATA 615
Qy	82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101
Db	616 TTGGCCCTGCCAGGAGGACTTATACCAATAGCTGGACAGATATGAATGGTGTCTATTTGCGT 675
Qy	102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyValAsn 121
Db	676 TCTACTGACGAAGAGATACCTTTTGAATCACTCCACTYCCCTTTTAAATTTGGTGGGAAC 735
Qy	122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141
Db	736 ATCCCRKGAAGGAATCTGGGGAAGAGTTAGCTATTGATCCAAACATAATAGATAATTA 795
Qy	142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161
Db	796 TACTTAGGAACAAGAGAGMATGGGTTATGGAAAAGTGAAGATATAGGAGTGTCTATGG 855
Qy	162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 181
Db	856 AAAAAAGTAAAGATTTCTTAACCCAGGACATATATGAAGATCCAAATCGCCCAAC 915
Qy	182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201
Db	916 GATTATTTGAACCATATAACAGTGTGTCTGGGTGATTTGATCTCTACAGTGTGTAGG 975
Qy	202 LeuGlyGlnAlaSerIleThrIlePheValGlyValAlaAspProAsnAsnProValPhe 221
Db	976 CTGGTGGAGGAACCAAGATAATTTATGTGGTGTGTGCAGATAAAACCAACAGCATCTAT 1035

Qy	222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241
Db	1036 TATACAAAAGACGGTGGTGCAGCATGCGCAAGCATTTACCAGGACAGCAACTGGCTACTT 1095
Qy	242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261
Db	1096 CCTCAGCGTGCAAAATTAAGTTCT---GATGGGATGTTATATATATATATATAGTAACT 1152
Qy	262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281
Db	1153 CAGGGCCATATATATGATGATGATGAGAAAGTATGAGATACATACATACACTAAGACCGAGAA 1212
Qy	282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301
Db	1213 TGGAAAATAATAAGTCCCAATGCTGCTCAAGATACAT---TATTTGGATATGGT 1263
Qy	302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321
Db	1264 GGATTTGGCTGTAGACGCCCAAGTCCGAAAGTAGTGGTAGCAGCTCTTAGCTCATGG 1323
Qy	322 TrpProAspThrIleIlePheAspSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341
Db	1324 TGGCAGATACCTATATTTGGAGAGTATCGATGTTGGGAAACGTTGGAAATGTTATTTGG 1383
Qy	342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361
Db	1384 GAATGGATGCTATCTTAAATAGAACATTCATATATACATGATATATTTCTGAGCTCCT 1443
Qy	362 TrpLeuThrPheGly---ValGlnProAsnProValProSerProLysLeuGlyTrp 380
Db	1444 TGGCTAAACTTTGGAAATCTATCTTACACACCTCGAAGTAAAGCCCAAACTTGGCTGG 1503
Qy	381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db	1504 ATGGTTGGAAACACTTGAATAAGACCAATTAATTCAGATARAATGCTTTATGGAAACAGGA 1563
Qy	401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
Db	1564 GCTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1623
Qy	421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440
Db	1624 AAAGTCAAGGTATAGGAATTTGAAGAACTTCAGTCCCAAGCACTGATAAGCCACCACTT 1683
Qy	441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
Db	1684 GGACCACTTTATTCAGTGCATCTGGAGATATAGCGGATTCAGACATGAAGACTTTGGAG 1743
Qy	461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db	1744 AAAGCACTAATTTGGACATATGTTTCACTAATATGTTGGGACTACTACTGATATGATTTT 1803
Qy	481 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db	1804 GCTGAGTTAAATCCAAATTTTATGTTTCTGTTAGGAATGTTGATAAACAGTGAATCCA 1863
Qy	501 AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGly---Ser 519
Db	1864 AATACTAACAAATTTGTTCTCATATGATGTTGGAAATCATGTTTTCAGGAAATACT 1923
Qy	520 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539
Db	1924 GAACCTCAAGGAACACAGCGAAGTGGTACAGTTGCGCTGCTGCGAGATGTTAGTGTCTGT 1983
Qy	540 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 559
Db	1984 GTTTGGGACCA-----AAAGGACGGAAGTATGCTATTTCTACAGATAACCGAAACAAA 2037
Qy	560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579
Db	2038 TGGGTGAATCGCAAAATGTAACCTTCAGAAAGTATTTGATATATTTCTGACAGAGTAAATCCA 2097







```

VKNPSTWCSGKTLVDRTSATVTVLVNHSVTSGLGDKVEKLPALKABASSANKTLEA
ANNAEKHMDAMWATNKINGINNAKEVVDLKKLGNKGNKTNFEPKQNFYKDKNLL
KDKGNLKPQDQLLIQQDQLETFEKNNTLAKNKYKNSRDLNKNYRDIKASLYNIPA
GGTLVNTEDVQSGFHFNPGEIKPIFEEIKVVLGGR"
gene
/sequence="CAC0924"
6345..6923
/sequence="CAC0924"
6345..6923
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/db_xref="GI:15023825"
/db_xref="GI:15023825"
translation="MIRFOKRMALFFLTISILICGDIIEKIIISFGIEFDDILCNLFFI
GLGIIGFRIYRWYKQIKYKTELKEDRLDKIEYIYDIERVGVWYNTAFYIDGVKC
KLFPOYCKNDNDYVYLEPIYKNGEKFDFAYADKNDYKYSIESTNEITECEBK
AEKKRKRLNEYVNSRMELAKKYEBEAKNM"
7003..8043
/sequence="CAC0925"
7003..8043
/sequence="CAC0925"
7003..8043
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/db_xref="GI:15023826"
/db_xref="GI:15023826"
translation="MENENVLRAQELNNRGLIFERNAGDEALKYFNKAIKEDENKFE
AVLNKAALCLANNKVDGAVSCYDKLIIVKYNKGAIVFGKGNVFFYKDNIKKALELVN
RSIYLGEMSEGVYINLALCMESIGELDDAIKWLNRALIINDQNIITLNKAAVLKIK
KPYEALCEKYNVLKIVDNEEVYHFKMLGEMGKIEKAPETISRGALIGEQMTLCY
DKAVLPEKQEFKALCEKELALLFDEGNVLLIKKGLLTYLKKIDEAKSTYDKLIE
FHNEMGVEFPAKANLQMLLGEHEDABEFFKEIIRLSGBESVLINSYVYRALNKKWA
KTRKQKRLIRKQ"
7998..8357
/sequence="CAC0926"
7998..8357
/sequence="CAC0926"
7998..8357
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/db_xref="GI:15023827"
/db_xref="GI:15023827"
translation="MGNKEEAKYKAEIKKYNFLIKYPIYDVYKLMFFKNCILRIED
YKAEELYEYIIDLNDALDEVLYLMRARNRISLNKYEAKEDLNTVIKLNPAVRKVIEL
DEOLKEFLSGKSSCYR"
8379..8675
/sequence="CAC0927"
8379..8675
/sequence="CAC0927"
8379..8675
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF-2 family"
/db_xref="GI:15023828"
/db_xref="GI:15023828"
translation="MNSDEIRIDEKACDDVDIKLAVGKVKETYNLIHLNSRHIENG
IKGSADVSIITEYKLVKVKENQDNNDVAANNVIVSSFDAEDKWKTKFYR"
8713..9012
/sequence="CAC0928"
8713..9012
/sequence="CAC0928"
8713..9012
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B.subtilis"
/db_xref="GI:15023829"
/db_xref="GI:15023829"
translation="MVDRCIDIEIDASIFETIKVYSDSKVRLHDIICDLENEILKMD
TWEGDAKKEFDSTFPGFYSAMKDCDMLBELIKELKLVKTSFESLDTMKNLDDK"
9712..10335
/sequence="CAC0929"
9712..10335
/sequence="CAC0929"
9712..10335
/codon_start=1

```

```

/transl_table=11
/product="SAM-dependent methyltransferase"
/protein_id="AAK78905.1"
/db_xref="GI:15023830"
/db_xref="GI:15023830"
translation="WGSQVDFDNTAAASNNVIRSEYFEERLKYKVLKSKVSKDKITADP
CGGTGPFISLALASDANWVGLDNNMLRELKSHDKLNNIYLKLSLTSVTLFDE
SIDAVFINMALHHIKDAEIKEMIRVLKDKGTVVISDVSEHNGWAEEMFDEWLGF
SNEQIDMWLKNVGFKNIDIENTDLCKGTSRGEHTETSIFIAKAA"
10353..11498
/sequence="CAC0930"
10353..11498
/sequence="CAC0930"
10353..11498
/codon_start=1
/transl_table=11

```

Alignment Scores:	1.34e-93	Length:	12732
Pred. No.:	2009.00	Matches:	377
Score:	65.41%	Conservative:	107
Percent Similarity:	50.95%	Mismatches:	240
Best Local Similarity:	49.78%	Indels:	16
Query Match:	1	Gaps:	12
DB:			

```

US-09-917-376-3 (1-740) x AB007608 (1-12732)
Qy 3 ThrGlnProTyrThrTpsSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIle 22
Db 181 AGCCAGGATATAGTGGGATATGCAAGATAGTGGTGGTATATGTTCCAGCAGTT 240
Qy 23 ValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyr 42
Db 241 ATATTTAATAAACAGAAAAAGATTATATATGCTGCTACAGATATGGGAGGAGCATAT 300
Qy 43 ArgTrpAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTrpValGlyTrpAsnAsn 62
Db 301 AGATGGGACACAGGCTAATAAAGTGGATTCCTAATCTGAT-----GGATTTCTGAT 354
Qy 63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTyr 82
Db 355 TGGCAATGCTTGGTGGGAAAGTATACCACTGACCAATAGACACTAATCGTGTAT 414
Qy 83 AlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArgSer 102
Db 415 ATACGACAGGCTTTGTACACAAATGATGGCAAGATGAAATGATATATCTTAGTTCA 474
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMet 122
Db 475 CAGGATAAAGGTAACACATGGAAAAAGATACCAACTTCCTTAAAGTAGGAGGTAATATG 534
Qy 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr 142
Db 535 CCGGAAAGAAATATGGGAAAGAAAGCTTCAATAGATCTCTAATAGATGATAAAATCTTTAT 594
Qy 143 PheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSer 162
Db 595 CTTGGAGCTGAAGTGGTAATGGACTTTTGGAAAGTGAAGACTATGGACAAACATGGTCT 654
Qy 163 GlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182
Db 655 AAGGTTGATAATTTCCCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
Qy 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202
Db 709 TATACAGCAGATAAAGTTGGAGTAGTTGGGAAACCTTTTGATCTCTCAACAGGACAAAA 768
Qy 203 GlyGlnAlaSerLysThrIlePheValGlyValAlaAsp---ProAsnAsnProValPhe 221
Db 769 GGAAGCCCACTCAACTATGATGTTGGTGCAGCAGATAGACAGGAAACAATATTAT 828
Qy 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241
Db 829 GTTACTAATGATGAGGAAAGACATGAGTGCAGTTAAAGGACGACGCAAAAGGATATCTT 888

```

Qy	242	ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr	261
Db	889	CCACATCATGGTATACTT---GCTTCCGATGGTATGCTTTATATATCATATAGCAATACT	945
Qy	262	GlyGlyProTyrAspGlySerSerGlyAspValTrrpLysPheSerValThrSerGlyThr	281
Db	946	TGGCGGACCATATGACGGCTCAGATGGACAAAGTTTGGAAGTATATACCAAAAACAGAGAA	1005
Qy	282	TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer	301
Db	1006	TGGCAACAATACTCCACCTGCAGTAGGTGACACTAAATCT-----GGTTTGGGA	1056
Qy	302	GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp	321
Db	1057	GGAATATCAGTAGATGCACAAAAATCCAAATAATAGTTGTAGTTGCCAATTTAAATATAGATGG	1116
Qy	322	TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrrpThrArgIleTrrp	341
Db	1117	TGGCCGGATGAAGAAATATATACAAAGTACAGATGTCAGGTGAAGCTTGGAACCTATATGG	1176
Qy	342	AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro	361
Db	1177	GATGGGAATGGATATCCAAACCGTACTCTTGGGATATAACTTAGATTACTCAGCACAACT	1236
Qy	362	TrpLeuThrPheGly---ValGlnProAsnProProValProSerProLysLeuGlyTrrp	380
Db	1237	TGGCTTGATTGGGGTAAGACCGGGTAACACCAACCCAGATCCATTAGTTAAATTAGGCTGG	1296
Qy	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly	400
Db	1297	ATGATGGGAGACTTAGAAAAATGTAGTCAATTTAACTCGGATAGATGTTCTACGGTACAGGT	1356
Qy	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrrpAspSerGlyGlyGlnIleHisIle	420
Db	1357	GCAACTCTTTATGGAACCTGATGATTTGACATAATGGGATAAAGTAAGATGATGACATC	1416
Qy	421	AlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer	440
Db	1417	TCTGTAAGGCAAAATGGAAATGAAGAATGTGCAGTAAATGATGTTGTAGTACCACAAAA	1476
Qy	441	GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
Db	1477	GGCCGCAACATTTTAAAGCGCTGTGGAGATGACTCGGATTTTATCATGATGACATAACT	1536
Qy	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
Db	1537	AAAGTTCCAAAGTAAATGATGACACACCACTTAATTTAGTGTACACCAACGATTTGATTAT	1596
Qy	481	AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro	500
Db	1597	GCAGAAAGTGTCTCTAACTTTGTTGTGTGTTGGAAATGTTGATAGTAAAAATCAA	1656
Qy	501	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrrpPheGln---GlySer	519
Db	1657	CAAGATAAAGATTGTGGAAATTTCTACGATGGTGGTAAGAACTGGTTCTCAGCTGGAAGC	1716
Qy	520	GluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPhe	539
Db	1717	AATATATACGAGGATTTATAAAGCAGGAACTGTAGCAGCTGGTGTGATGCTAAAAACTATA	1776
Qy	540	ValTrrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer	559
Db	1777	GTATGGTCTCCA---GAGGAAAGTGTCTAATGTCAGCATATAGTCACAGATAATGTATAAAA	1833
Qy	560	TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro	579
Db	1834	TGGACACCGTGTCTGCTCTCCACAGGGAGCAAGGTTAGATCAGACCGTGTAAATCCA	1893
Qy	580	LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr	599
Db	1894	AAGAAATTCATGGTTTCTTAAATGTGTAATTTCTATATTAGTACTGATGAGGTGCAACA	1953

Qy	600	Ph	---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis	618
Db	1954	TTTACTCAAAAGTTTCACAAACAGGCTCTTCTACAAAAGGCAAA---GGAATC---	TTTAAAG	2007
Qy	619	AlaValProGlyLySgGluGlyAspLeuTyrPheAla---	AlaSerSerGlyLeuTyrHis	637
Db	2008	ACTGTTATAGTCTATGAAGGTGATATTGGATTGCTGGAGGAAAGACGATTAIGCAT	2067	
Qy	638	SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal	657	
Db	2068	TCACTGATCTTGGTGCACATTCACAAAAGTATCTGCTGTGATGCTCTGTATACAGTA	2127	
Qy	658	GlyPheGlyLySgSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle	677	
Db	2128	GGTCTTGAAAAATCCAAAACAGATGATGGATATCCAGCTATATACATGGATGCTACAA	2187	
Qy	678	GlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTyrValLeuIleAsn	697	
Db	2188	GATGGAAACAGCGGAATATTCCGTTTCAGATGATGAAGTGCACCTGGTAAAGATAAAT	2247	
Qy	698	AspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeu	717	
Db	2248	GATGATGCTCACCATTATGTTCCCGACTACTGCATTAAGTGGAGATCCTAACAAATAT	2307	
Qy	718	ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla	737	
Db	2308	GGACGAGTTTTTGTGTGAACCTAATGGAAGAGGTATTGTTTATGGAGACATAGATGGT	2367	
RESULT 7				
AJ585344		2950 bp	DNA	linear
LOCUS				BCT 11-OCT-2003
DEFINITION				Clostridium thermocellum xghA gene for beta-1,4-xyloglucan
ACCESSION				hydrolyase, strain F7.
VERSION				AJ585344.1 GI:37651952
KEYWORDS				beta-1,4-xyloglucan hydrolase; xghA gene.
SOURCE				Clostridium thermocellum
ORGANISM				Clostridium thermocellum
				Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
				Clostridium.
REFERENCE				1
AUTHORS				Zverlov, V.V.
TITLE				New cellulosome components of C. thermocellum
JOURNAL				Unpublished
REFERENCE				2 (bases 1 to 2950)
AUTHORS				Zverlov, V.V.
TITLE				Direct Submission
JOURNAL				Submitted (09-OCT-2003) Zverlov V.V., Fachgruppe f. Mikrobielle
				Mikrobiologie, TU Muenchen, Am Hochanger 4, Freising, 85350,
				GERMANY
FEATURES				Location/Qualifiers
source				1..2950
				/organism="Clostridium thermocellum"
				/mol_type="genomic DNA"
				/strain="F7"
				/db_xref="taxon:1515"
				346..2887
gene				/gene="xghA"
				346..350
RBS				/gene="xghA"
				359..2887
CDS				/gene="xghA"
				/note="belongs to GH74 and contains a cellulosomal
				dockerin module"
				/codon_start=1
				/evidence=experimental
				/transl_table=11
				/product="beta-1,4-xyloglucan hydrolase"
				/protein_id="CAE51306.1"
				/db_xref="GI:37651953"
				/translation="MVKKFTSKIKAAVFAVAATAIFGPAISQAVTSVPYKWDNVV
				IGGGGPMPIGVNETEKLIYARIDGIGAYRWDPSPTETWIPLLDHFQMDSEYSYGV
				SIATDPVDPNRVITVAGMYTNDWLNNMGAILRSTDRGETWEKTLIPFKMGNGMFRGM

GERLAI DPNDNRRLYLGTRCNGNLRSTDYGTWSKVESFPNCTIYIDBNFDYTDKI  
IGVVWVPDKSSDGNPTKILVGVADKNESIYSTDGGVTWAVPGQKGLLPHHG  
VLASNGMLYITGDTCPYDNGNGKNESTIGEWIDITPIYPSDNRKPCFAGLA  
VDRQDITIMVTSNMAWPDXYIFRSTDGATKQNIWGMTPYRILTHEYIDISAAFP  
LDMQTEKOLPEINLPGKLMWIGIDIPDPSDRMWTGTATIGYCDNLTDWRGCKYKI  
EVAKATIEEACVADLVSPEEGLAVSVAGDLGVFHHDLKVGPKMHVSVYSSTGDI  
YAEALPHNMAVAKADLVYDKPKISVSPDQGRNVRPQNPENASVGSVGVAVAKSV  
IWTENASPAVTTDNGNSWKVCTNLGMGAVLASDRVNGKIFYFYNGKFIYSTDGGIT  
PFTDPAQPKSVNKIYTKGIDNVGLFRRSDAGGLKWTWRINDDEHYGAVTAPIYTDGRPV  
FGKAAQPDQYVNIYATKIDVGLVGEPSADEFTPPQVDRKGLVGLDNGRINSTD  
YGRVYIANTGRGIVYGEPSADEFTPPQVDRKGLVGLDNGRINSTDLMKRYILK  
SIEDLPVEDDILWAADINDEGKINSIDTYTYPVRLGLQLOAIPELPPKK"

## ORIGIN

Alignment Scores:

Alignment Scores:		
Pred. No.:	6.26e-93	Length: 2950
Score:	1982.00	Matches: 387
Percent Similarity:	65.81%	Conservative: 100
Best Local Similarity:	52.30%	Mismatches: 229
Query Match:	49.11%	Indels: 24
DB:	1	Gaps: 14

US-09-917-376-3 (1-740) X AJ585344 (1-2950)

Qy	2	ThrThrGlnProTyrThrTrpSerAsnValAlaLile---	GlyGlyGlyGlyPheValAsp	20
Db	458	ACCAAGCGTGCTTTACAAATGGGACAAACGCTGCTAAATCGGCGAGGCGGAGGATTTATGCGG	517	
Qy	21	GlyIleValPheAsnGluGlyAlaProGlyLylleuTyrValAArgThrAspIleGlyGly	40	
Db	518	GGTATAGTTTTTAATGAACCGAAAAAGGATTTGATTTATGACGCTGCCGATATCGGAGGA	577	
Qy	41	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp	60	
Db	578	CGGTACCGGTGGGATCCTTCACCGAGACATGGATTCCGTGCTGGACCATTTCCAAATG	637	
Qy	61	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	80	
Db	638	GATGAGTACAGTTATTATCCGAGTGCMAAGATTTCGAACCGACCCCTGTGGATCCGAACCGT	697	
Qy	81	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu	100	
Db	698	GTTTACACATAGTTCGAGGTATGTATACCAACGATTGGCTTCCTTAATATGGGACCAATCTT	757	
Qy	101	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPropHelysLeuGlyGly	120	
Db	758	CGCTCAACGACGAGGGAGAAACATGGGAAAAACCATACTGCTCTTCAAGATGGGCGGA	817	
Qy	121	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	140	
Db	818	AACATGCCGGGAAGATCCATGGGAGAACGTCTTGGATCGACCCGGAATGCACAAACAGGATT	877	
Qy	141	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr	160	
Db	878	CTTTATCTTCGAAACACGATCGGAAACCGAGCTTTGGAGAAGTACCGACTACGGTGTAA	937	
Qy	161	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	180	
Db	938	TGTTCCAAAGGTGAAGTTTCCCAATCCCGAACTTACATTTATGACCCG	991	
Qy	181	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer	200	
Db	992	TTTGATTATACCAAGACATATTGGAGTAGTCTGGGTGTTTTTGAACAGACGACAGT	1051	
Qy	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220	
Db	1052	ACACCGGGCAACCTACCAAGACTATATATGTTGGTGTGGCTGATAAAAAACGAAAGTATT	1111	
Qy	221	PheTrpSerArgPheGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe	240	
Db	1112	TACCGCATACGACGGGGGTGCACCTGGAAAGCAGTTCCGGCAACACCTTAAGGAGCTA	1171	
Qy	241	IleProHilysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260	

Db	1172	CTTCCCTCACACGGGGTTTG--GCATCCAAACGGGAATGTTGTATATAA	CTTATGGTGAT	1228
Qy	261	ThrGlyGlyProTyrAspGlySer--SerGlyAspValTyrIlePheSerValThrSer	279	
Db	1229	ACTGCGGTCTTATGACGCGCAACGGAAGCTGAGTTTGGAAAGTCAATACACGTACA	1288	
Qy	280	GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly	299	
Db	1289	GGGGAATGGATAGATATATCACCCGATACCTTATTAACAGCAGTGTGACAAT--CGTTTCTGCG	1345	
Qy	300	TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle	319	
Db	1346	TTTGCAGGACTTCAGTGGACAGCGAATCTCTGACATTATATGTTAACTTCCATGAAC	1405	
Qy	320	SerTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg	339	
Db	1406	CGGTGGTGGCGGATGATATATTTTCCGACGACTGACGCGGAGGACTCATGGAAGAAT	1465	
Qy	340	IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla	359	
Db	1466	ATCTGGGAATGGGAATGTATCTCGAACGTATCTGCATTATGAAATAGATATTTCCGCA	1525	
Qy	360	GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly	379	
Db	1526	GCACCGTGGCTGATTTGGGAATCTGAGAAACAGCTGCCGGAATCAATCCGAAATCGGA	1585	
Qy	380	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr	399	
Db	1586	TGGATGATAGGTGACATAGATGTGACCCGTAAATTCGACCGCATGATGTATGTATACC	1645	
Qy	400	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis	419	
Db	1646	GGTGCAACTATCTATGTTGTGCAATCTTACTGACTGGGACAGAGCGCAAAATAAA	1705	
Qy	420	IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPro	439	
Db	1706	ATCGAGTAAAGCTACCGGAATAGAGAATGTGCGGTATTAGACCTGGTAAGCCGCGCG	1765	
Qy	440	SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal	459	
Db	1766	GAGGTGCACCGCTGTGAAGTGAGTGGGACCTGTCTCGGTTTGTTCATGATGACCTG	1825	
Qy	460	ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp	479	
Db	1826	AAAGTTGGTCCGAAAAAATGCAGT--CCTTCTTATTTCTTCAGGTA	1882	
Qy	480	TyrAlaGluLeuAsnPro-----SerIleIleValArgAlaGlySerPheAspPro	496	
Db	1883	TATGCGGAGCTTGTTCGAACTTTATGGCATTTGGTTCGCAAGGCTGATTTGTATGATGA	1942	
Qy	497	SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe	516	
Db	1943	-----AAGAAGATTTCTTCTTATACGCGAGAGAAATTTGGTTC	1984	
Qy	517	Gln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAsp	535	
Db	1985	CAGCCACCTAATGAGCACCAAACTCGTAGGCGCGGTTCCGTTGCGTTGCAGCGCAT	2044	
Qy	536	GlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGly	555	
Db	2045	GCAAAATCAGTTATTGGACACCG--GAAATGCAAGTCCTGCAGT---ACAAACGAC	2098	
Qy	556	PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAsp	575	
Db	2099	AACGGAAACTCATGGAAGATTGTGACAAATCTTGGTATGGGTGGGTTGGCATCCGAC	2158	
Qy	576	ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp	595	
Db	2159	CGTGTGACCGTAAATAATCTACGCAATTCATAAACGGCAAAATCTATATATAGACGAC	2218	
Qy	596	GlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAlaValGly	614	

Db 2219 GTGGGATTAACCTTTACCGATACAAAGGACCGAGCTTCCCAAGTCGGTTAAC----- 2272

QY 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAa---SerSer 633

Db 2273 ---AAGATAAAACCCGTACCGGCAAGGAAGACATGATGGCTTCTGCAAGAGAAGGC 2329

QY 634 GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSer 653

Db 2330 GGATTTGGAGGTCCACTGACGGTGGATATACGTTTGAGAAACTCTCCAAATGTTGACACA 2389

QY 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673

Db 2390 GCTATGTGGTACGCTTCGGAAGGACGACCGGACAGGATACATGCGGATTTACATT 2449

QY 674 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 693

Db 2450 ACCGTAAATTCACATGTTTAGATCTTCGTTCCGATGATGCCGCAAGACATGG 2509

QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713

Db 2510 GTGGGTATCAACGACGACGACGATATGGCTGTTGATCTACTGCAATAACAGGTGAC 2569

QY 714 HisAlaAsnLeuArgValTrpIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733

Db 2570 CCGAGAGTATACGAGCTGTATATATTGCCCAACGGAAGGATATTGTTTACGGCGAA 2629

RESULT 8

ABI292929

LOCUS Agaricus bisporus mRNA for CEL6 protein. linear PLN 08-AUG-2000

DEFINITION Agaricus bisporus mRNA for CEL6 protein.

ACCESSION AJ292929

VERSION AJ292929.1 GI:9755896

KEYWORDS cel6 gene.

SOURCE Agaricus bisporus

ORGANISM Agaricus bisporus

Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Agaricaceae; Agaricus.

REFERENCE 1

AUTHORS Morales-Almora, P. and Thurston, C.F.

TITLE Molecular analysis of the cellulolytic genes in Agaricus bisporus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2579)

AUTHORS Morales-Almora, P.

TITLE Direct Submission

JOURNAL Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's College London, 150 Stamford Street, London, SE1 8WA, UNITED KINGDOM

## FEATURES

source

1..2579

/organism="Agaricus bisporus"

/mol\_type="mRNA"

/strain="D649"

/db\_xref="taxon:5341"

/clone="PMRT152"

/dev\_stage="mycelium"

/country="United Kingdom"

1..2579

/gene="cel6"

28..2448

/gene="cel6"

/function="putative cellulase"

/codon\_start=1

/product="CEL6 protein"

/protein\_id="CAC02964.1"

/db\_xref="GI:9755897"

/db\_xref="GOA:Q9P4T8"

/db\_xref="SPRENEL:Q9P4T8"

/translation="MPSRVKFWLITLIGTFLSYVGOAASQSYRNQNVKIGGGGFPVPG IVFNPSKGLAYARTDIGAYKLNADDTWTFLLDFADDSRWNTYGVDAIATDVQPNR LYLATGLYNSWDPNSNGHILISSDYDKTFDAAPFLKLGNNPGRMGRLVVDPNLN SILYFCARSGNLWKSNGSRWSKVTSTFDTPGSPDSDTGLNSDKIGIAWTVID KASGSGSATPRTFVGVANKGNSIYISTNGSGSSAVAGQPTSLPHKGLVLSERA LITYSDGGPGDGTSGAVYKYSIDITGAWTDITPVSGLDLTFPGGLALDITQNGTVM

## gene

## CDS

VAALNSWPDQIIFRSTNGGASWTPLWDVAVPTLNKYYSYNAALAPWIGPNIVDTP  
GNIQIWMESLSIDPDPFNSHLYGTGETIYGSRLDLKWDSTAHNTYIISLADGVBETS  
VQALISPSGPPVSAIGDVGCFVHQSTRAPSOFTNPKWTSADIILFAGNPNVNV  
RIGDTSGKQVAISSDYVTNQHFCAPDNVCGKVAISADADIILPWTNGNGVMV  
SRQATFNIFYGASGKTFYVSTDKTKTFSAGSLGATSVFDITVHPVSGDIDWASTD  
KGLFHSSTDGATFSAISIGITOAQWALGAPRSTGYPAVPAANAATGGEVAYFRSDRG  
VNWVKINDAAHFGGAASANCMAADPRVYRTVINGRGIFYGDVAGSAPPPTSSATS  
TTSYVTTTTRSTTTTTRSTTTTSSGNSGVSGFGCGGQGGTGGTTCVTGFOCVAQNP  
FYSOCIP"

28..96

/gene="cel6"

97..2445

/gene="cel6"

/product="CEL6"

sig\_peptide

mat\_peptide

ORIGIN

Alignment Scores:

Pred. NO.: 2,85e-81 Length: 2579

Score: 1753.50 Matches: 341

Percent Similarity: 60.97% Conservativity: 112

Best Local Similarity: 45.90% Mismatches: 249

Query Match: 43.45% Indels: 41

DB: 8 Gaps: 10

US-09-917-376-3 (1-740) x ABI292929 (1-2579)

QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheVal 19

Db 94 GCAGCTTCTCAGTCTTATAGGTGGCAGATGTCAAGATCGCGGGTGGTGGTGGTGGTTC 153

QY 20 AspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 39

Db 154 CCGGCGATTGCTTCAATCCTTCCGAAAGGAGCTGGCGTATGCTGCACACTGACATCGA 213

QY 40 GlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly 59

Db 214 GGAGCTACAGCTGAAT---GCTGATGATACTTGGACGCTCTACTTGAATTCGCTGAT 270

QY 60 TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn 79

Db 271 GATTCTCGTGGGAATATTGGGGCGCTCGATGCTTTCGAACATGATCCCGTACCAACCAAC 330

QY 80 LysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle 99

Db 331 CGTCTTTACCTAGCAGCTGGGTGTATACCACTCTGTGGAGCCCAAGTAAAGCCATATT 390

QY 100 LeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 119

Db 391 CTCATATCAAGTGATTTATGTTAAACATTCGATGCCGCGCTTTGCCATTCAAGCTCGGT 450

QY 120 GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsn 139

Db 451 GGTAAATATGCTCGGCGTGGAAATGGAGAGCGCTCTCGTCTCGATCCCAATCTGAACAGC 510

QY 140 IleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAla 159

Db 511 ATCCTTTATTTTGGCGCTCGCAGTGTAAACGGTCTCTGGAAGTCAACAATTCGGGTCCG 570

QY 160 ThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAsp 179

Db 571 AGCTGGAGCAAGGTAAACAAGCTTCACTGACACTGGGTCTTTCTGTTCCCGCATCCAT 630

QY 180 ThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 199

Db 631 AGCACCGGCTTGAATTTCTGATAAGATTGGGATTCGATCGATGGGTTCACCATCAAGCGTCA 690

QY 200 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsn 218

Db 691 GGCTTCTGCTCGCTACCTCCCGTATATTTTGTGGCGTGTGGCAACAGGCGCTCAAC 750

QY 219 ProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr 238

Db 751 AGCATTTATATATCTACGAATGGCGGTCTTCTTGGAGTGGCGTCCGAGGTCAACCAACT 810

239 GlyPheIleProHisIysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr 258  
 811 AGCTTCTCTGCCACATAAGGCGGTTCTATCCCTCTCTGAGAGGCGACATTATATCACCTAC 870  
 259 SerAsnThrGlyClyProTyrAspGlySerSerGlyAspValTyrIlePheSerValThr 278  
 871 TCCGATGGTGGCGGCTCTTACGATGCGACGAGCGGTGCGATTTACAGATGACGATTCGAC 930  
 279 SerGlyThrTyrThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 298  
 931 ACTGGCGCATGGAGTACATACACCCGCTGAGCGGAGCGATCTC-----TTCCTTT 981  
 299 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 318  
 982 GGTTCGGAGTCTAGCCTTAGATACACAGAAACCGGACTGTCTATGTTGCGAGCCCTC 1041  
 319 IleSerTyrTyrProAspThrIleIlePheArgSerThrAspGlyAlaThrTyrThr 338  
 1042 AACTCGTGGTGGCGCTGATGTCAGATATTCGCTCGACCAACGGGGTGCATCTTCGACT 1101  
 339 ArgIleTyrAspThrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 358  
 1102 CCCTGTGGGACTGGGAGTCTACCCGACTCTCAACAAATACATGTTATACGCTGCC 1161  
 359 AlaGluProTyrThrPheGlyValGlnProAsnProProValProSerPro----- 376  
 1162 CTGGCGCGGTGG-----ATAGGACCCACACATCGTTGATGTCACCCCGGAAT 1209  
 377 ---LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMet 395  
 1210 CTACAAATCGGTGGATGATGGAATCACTATCCATTCATTCGACTCGAATCATTTGG 1269  
 396 LeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSerGly 415  
 1270 CTCTACGCGACTCGAGAAATCTATCTATGGAAGTGGTGCACCTCTCAATGGGACAGTGCA 1329  
 416 GlyGlnIleHisIleAlaProMetValIysGlyLeuGluThrAlaValAsnAspLeu 435  
 1330 CACACGTCACTACATCAAGTCGTCGCGAGTGGTTCGAGGAAACCTCGTCCAGCCCTT 1389  
 436 IleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThr 455  
 1390 ATTTCACCTCTTCGCGACCGCTCTGTTCTGCTATTGCTGACGTTGGCGGTTTGTGTC 1449  
 456 HisAlaAspValThrAlaValPheThrIlePheThrSerProValPheThrThrGly 475  
 1450 CACCAATCACTTACCCGTCGACCGCCGCGTTCACCAATCCCAAGTGGTGCACCTCG 1509  
 476 ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAsp 495  
 1510 GCTGACATCATTTTCGTCGCAATACCTAGTAAATGCTGCTTCGATGTTGCTGTTGAT 1569  
 496 ProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTyr 515  
 1570 AGTACTAGT-----GGCAAAACAGGTGCGCATCTCTCCGACTATGGCGTTACTTGG 1620  
 516 PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAsp 535  
 1621 AATCAGCATTTTGGAGCTCCTGATAATGTTCAAGGCGGAAAGGTTCGCATCTCCGCTGAT 1680  
 536 GlySerArgPheValTyrAlaProGlyAspProGlyGlnProValValTyrAlaValGly 555  
 1681 GCCGACATCATCTCTGGCGGTCAAAATGGTAACGGT----- 1716  
 556 PheGlyAsnSerTyrAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAsp 575  
 1717 -----GTCATGGTGCAGGAAC 1734  
 576 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 595  
 1735 CAGGCCACCTTCAATATTTCTACGGAGCCTCGGCAAAACATTTACAGTTTCCACAGAC 1794

596 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 615  
 1795 AACGGCAGACATTTTCAGCCCATGGTAGCTTACGCTTACGCTTCTGTCTTCGACATA 1854  
 616 MetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeuAlaAlaSerSerGlyLeu 635  
 1855 ACTGTCCAC-----CCGAGTGTGAGCGGTGATATCTGGCCCTCCACATGACAGGCGCTA 1908  
 636 TyrHisSerThrAsnGlySerSerTyrPheSerAlaIleThrGlyValSerSerAlaVal 655  
 1909 TTCCATTTCTACGACGCGGTGCGACTTTTAGTGTATCTCTGTTATCACACAGGCGCTG 1968  
 656 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGly 675  
 1969 GGTGTGCTCTGTGGTGCACCTAGATCTACAGGAGGATATCTGCTGTTTTTTCGTCGCGCC 2028  
 676 ThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTyrValLeu 695  
 2029 AATTACGCGCGGAGGTAGCTTACTTTAGTGTCTGATGACCGGGAGTCAACTGGGTAAAG 2088  
 696 IleAsnAspAspGlnHisGlnTyrGly---AsnTyrGlyGlnAlaIleThrGlyAspHis 714  
 2089 ATCAATGACGCGGCACATCGCTTGGTCCGCTAGTCTTAAGTGTGCTGATGCCA 2148  
 715 AlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 734  
 2149 CGTGTATACGCGACGCGTTTATATCGAACCAATGTCGCGGTATCTTCTATGCTGATGT 2208  
 735 GlyGlyAla 737  
 2209 GCTGGTTCC 2217  
 RESULT 9  
 LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998  
 DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.  
 ACCESSION AB015511  
 VERSION AB015511.1 GI:3242654  
 KEYWORDS Avicelase III.  
 SOURCE Aspergillus aculeatus  
 ORGANISM Aspergillus aculeatus  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 REFERENCE 1 (bases 1 to 2745)  
 AUTHORS Arai, M., Takada, G., Kawaguchi, T. and Sumitani, J.  
 TITLE Avicelase III from Aspergillus aculeatus  
 JOURNAL Published Only in Database (1998)  
 REFERENCE 2 (bases 1 to 2745)  
 AUTHORS Arai, M., Takada, G., Kawaguchi, T. and Sumitani, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-1998) Motoko Arai, Osaka Prefecture University,  
 College of Agriculture; Gakuencho 1-1, Sakai, Osaka 599-8531, Japan  
 (E-mail: motoko@biochem.osakafu-u.ac.jp, Tel:81-722-54-9465,  
 Fax:81-722-54-9465)  
 FEATURES  
 source  
 1. 2745  
 /organism="Aspergillus aculeatus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:5053"  
 gene  
 1. 2745  
 /gene="aviIII"  
 CDS  
 8. 2578  
 /gene="aviIII"  
 /codon\_start=1  
 /product="Avicelase III"  
 /protein\_id="BA029031.1"  
 /db\_xref="GI:3242655"  
 /translation="MARSSALLCAALIKGLADAAASOAYTWKVVVTGGGGFTPGIV  
 FNPASKGVAYARTDIGNAYRLNSDDTWTPLMDVGNDDTWHWDGIDALADTDPVDTDRVY  
 VAVGMYNEDPNVSGSLRSTDGDTWTETKLPFKVGGNMPGSGRGLERLAVDPNKNKI  
 LYFGARGHGLKNSDTGRTSTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  
 SSGSATPRI FVGVADAGKSVFKSDAGATMAWVSGEPQYGLPHKGLSPSEKTLVSI  
 YANGAGPYDGTNGTGVHKNYITSGVWTDISPTSLASTYYGYGLSLVDLQVPGTLMVAAL

NCWPDLEIFRSTDSGATWSPIEWNGYPSINYYSYSDISNAPWIQDTTSDQPPVRV  
 GNVVEALADPFDSNHWLYGTGLTYGGHDLTNWDSKHNTVTKSLAVGIEMVGLGI  
 TPFGKALLSAGVDDGFGYHSDLDAAPOAYHTPTTYTNGIDYAGNKPNIYVRSGAS  
 DDYPTLALLSNFGSTWADVAASTSTGTGAVALSADGDTVLLMSSTGSLVSKSQGL  
 TAVSLPSGAVIASDKSDNTVFGGSAGAI VVSNTATSTTKTVSLGSSITVNAIRAH  
 PSIAGDVASTDGLHSTDYDSTPTQIGSVTAGWSPFGKASSTGSYVVIYGFPII  
 DGAAGLFKSDAGDNTWQVI SDASHFGSGSANVNGDLQTYGRVFRGHERPGLHLROS  
 QREPARGHGDGDTTSTKSTSTVSTLTKTTSSASTSSSTVTKTTSSTSSSTSKAS  
 STTTTKTTTSTTSSGTTATATAYACGGNGWTGATVCTGTGYCTVSNFYSCQVPS  
 "

## ORIGIN

## Alignment Scores:

Pred. No.: 1,26e-79 Length: 2745  
 Score: 1722.00 Matches: 353  
 Percent Similarity: 62.27% Conservative: 114  
 Best Local Similarity: 47.07% Mismatches: 250  
 Query Match: 42.67% Indels: 33  
 DB: 8 Gaps: 17

US-09-917-376-3 (1-740) x AB015511 (1-2745)

Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaLalle---GlyGlyGlyPheVal 19  
 Db |||||  
 Qy 68 GCGGCTCGCAGGCGCTACACGTGAAGAACGTCGTCACGGGTGGCGGAGGAGGTTTCACC 127  
 Db |||||  
 Qy 20 AspGlyIleValPheAnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 39  
 Db |||||  
 Qy 128 CCCGGTATGCTCTCAATCCCTCCGCCAAGGCGTGGCGTACGGCGGCACGGATTCGGA 187  
 Db |||||  
 Qy 40 GlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly 59  
 Db |||||  
 Qy 188 GCGCGTACCGTCTGAAC---TCGGACACACGTGGACCCCGTGTGATGCGTGGTGGGC 244  
 Db |||||  
 Qy 60 -----TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIle 76  
 Db |||||  
 Qy 245 AATGATACCTGGCAGTACTGG-----GGCATCGATGCCCTTGGCAGCGATCCGGTC 295  
 Db |||||  
 Qy 77 AsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAsp 96  
 Db |||||  
 Qy 296 GACACGATCGTGTGACGTGGCGGTGGCGATGATACCAATGAATGGACCCCAATGTG 355  
 Db |||||  
 Qy 97 GlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleTrpLeuProPhe 116  
 Db |||||  
 Qy 356 GGATCGATCTCTGCTCGACGACACAGGTGATACCTGGACTGAGACGACTGCCGTTC 415  
 Db |||||  
 Qy 117 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 136  
 Db |||||  
 Qy 416 AAGTGTGGTGGGAATATGCCCGGTGGCGAATGGGTGAGAGATTGGCGGTGACCCGAAC 475  
 Db |||||  
 Qy 137 AsnAspAsnIleLeuTyrPheGlyAlaProSerSerGlyLysGlyLeuTrpArgSerThrAsp 156  
 Db |||||  
 Qy 476 AAGAATAGCATCTCTACTTTGGAGCGCGGAGTGGCGACGCGCTTGTGGAAGACACAGAT 535  
 Db |||||  
 Qy 157 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 176  
 Db |||||  
 Qy 536 TATGAGGACCTCGTGGATGCAATGTCCTCTACGTGGAGTGGAAAGTATTTCCAGAT 595  
 Db |||||  
 Qy 177 ProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAsp 196  
 Db |||||  
 Qy 596 TCGAGCTCGAC-----TACACTTCGACCCCTGTGGGTATTGCTTGGGTGACCTTTGAT 649  
 Db |||||  
 Qy 197 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216  
 Db |||||  
 Qy 650 TCGACCTCTGCTAGCTCGGGGTCTCGCACGCTAGGATCTTCGTGGAGTCCGCGATGCT 709  
 Db |||||  
 Qy 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 236  
 Db |||||  
 Qy 710 GCGAAGTCGCTCTCAAGTCGAGAGATCGGGGTGCTACTTGGGCTGGGTGCTCGGAGAG 769  
 Db |||||  
 Qy 237 ProThr---GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyr 255  
 Db |||||  
 Qy 770 CCTCAGTATGGCTTCTTCCGCCACAAAGGGGTCTTGTCCTCTGAGGAGACGCTGTAC 829  
 Db |||||

Qy 256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPhe 275  
 Db |||||  
 Qy 830 ATCTCTACCGGAATGGCGTCTCTTACGATGGCACAACGGAACAGTCCACAAGTAC 889  
 Db |||||  
 Qy 276 SerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsn 295  
 Db |||||  
 Qy 890 AACATCACAGCGGAGTGTGGACGATATAAGTCCC-----ACCTCCCTTGGGAG 940  
 Db |||||  
 Qy 296 AspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal 315  
 Db |||||  
 Qy 941 ACCTACTACGGCTACGGTGGTCTGCTGGTGGCTCCCAAGTCTCTGGCACTTGATGGG 1000  
 Db |||||  
 Qy 316 AlaThrGlnIleSerSerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAla 335  
 Db |||||  
 Qy 1001 GCTGCACCTCAACTGCTGCTGGCGGATGAGCTCATCTTCGTAGCACTGACTCTGGAGCG 1060  
 Db |||||  
 Qy 336 ThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeu 355  
 Db |||||  
 Qy 1061 ACCTGGTCCCTATTTTGGAGTGGAAACGGTTATCCGAGCATCAATTTACTACTACGCTAT 1120  
 Db |||||  
 Qy 356 AspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValPro 374  
 Db |||||  
 Qy 1121 GATATCAGCAACGCCCTTGGATCCAGGACAGCACTTCGACCGGACCAATTCGGGT--- 1177  
 Db |||||  
 Qy 375 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
 Db |||||  
 Qy 1178 -----CGCGTGGTGGATGGTGAAGCCCTTGCCATTTGACCCATTTGACTCCACACC 1231  
 Db |||||  
 Qy 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414  
 Db |||||  
 Qy 1232 TGGCTCTACGGCAGCGGCTGACATATATGGCGGACAGCACTGACCAATTTGGGACTCG 1291  
 Db |||||  
 Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
 Db |||||  
 Qy 1292 AAGCACAAATGTGACGGTCAAGTCACTGGCGGTGGTATTGAGAGAGATGGCAGTCTCGGA 1351  
 Db |||||  
 Qy 435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
 Db |||||  
 Qy 1352 CTGATACGGCGCGCGCGGACCGCGCTGTCTGTCTGCTTGGCGAGCAGCGCGGGTTC 1411  
 Db |||||  
 Qy 455 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 474  
 Db |||||  
 Qy 1412 TACCACCTCGGATCTGGACGCGCGCCAAACCAAGCGTACCACACGCGGACCTACGCGCAG 1471  
 Db |||||  
 Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 494  
 Db |||||  
 Qy 1472 ACCAACGCGCATCGACTACCGCGGCAACAGCCTTCCAAATCGTCCGCGTCCGAGTAGC 1531  
 Db |||||  
 Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsn 514  
 Db |||||  
 Qy 1532 GAC-----GACTATCCGACC-----CTGGCTCTCTGTCGAACTTCGGATCGACT 1576  
 Db |||||  
 Qy 515 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 534  
 Db |||||  
 Qy 1577 TGGTATGACAGACTACCGCGCTCAACACGACCGGACCGGAGCGGTGGCTCTTTCCGCC 1636  
 Db |||||  
 Qy 535 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 554  
 Db |||||  
 Qy 1637 GATGTTGACAGGTGCTTCTCATGTCTCCAGCACCTCGGGGCGGTGGTGTGGAAGTCCGAG 1696  
 Db |||||  
 Qy 555 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 574  
 Db |||||  
 Qy 1697 GGC-----ACCTTGACTGCGGTTCCTCTCTCCCGTCCGGGCGGTGTCATCGGCTCC 1747  
 Db |||||  
 Qy 575 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 594  
 Db |||||  
 Qy 1748 GACAAGTCCGACACACAGGTCTTCTATGGTGGATCCGCGGCGCAATCTACGTGTGGAAG 1807  
 Db |||||  
 Qy 595 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 614  
 Db |||||  
 Qy 1808 AACACAGCGACTCTCTTCCACCAAGACCGTGAGC---CTGGGGTCCAGCAGCAGCGCTCAAC 1864  
 Db |||||



```

Qy 615 valMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 634
Db 1865 GCCATC---CGCGGCATCCCTCCATCCCGCGCATCTCTGGCGTCGACGCAAGGGC 1921
Qy 635 LeuTyHisSerThrAnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSer 653
Db 1922 CTCTGGCATCCAGATTACGGTAGTACGTTTACGAGATCCGACGCGCGTCACCGCG 1981
Qy 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpTrpAlaValPheVal 673
Db 1982 GGCTGGAGCTTGGTTTGGCAAGCGCTCTCCACGGCGAGTACGTCGTCATCTACGGG 2041
Qy 674 valGlyThrIleGlyGlyValThrGlyAlaTyArgSerAspCysGlyThrTrp 693
Db 2042 TTCTTACCATTCCAGCTGCGCGGGCTGTTCAAGAGCAGGACGCGCACCACCTGG 2101
Qy 694 ValLeuIleAsnAspAspGlnHisGlnTyArgGlyAsnTrpGlyGlnAla-----IleThr 711
Db 2102 CAGGTGATCTCGGATCGTCGACGGATTGCGCTCG---GGTCCGCCAACGTCGTGAAC 2158
Qy 712 GlyAspHisAlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTy 731
Db 2159 GCGCACTGCAGACTACGCGCGGTGTTCTGTCGGCAGCAGCGCGCGGCATCTCTA 2218
Qy 731 rGlyAspIleGlyGlyAlaProSerGly 740
Db 2219 CGGAATCCCGAGCGGAGCTCGCGGGC 2246

RESULT 10
AY040839
LOCUS AY040839 3959 bp DNA linear PLN 01-APR-2002
DEFINITION Aspergillus niger endoglucanase C (eglC) gene, complete cds.
ACCESSION AY040839
VERSION AY040839.1 GI:19879405
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 3959)
AUTHORS Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de
Graaff,L.H.
TITLE EglC, a new endoglucanase from Aspergillus niger with major
activity towards xyloglucan
JOURNAL Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
MEDLINE 21914059
PUBMED 11918668
REFERENCE 2 (bases 1 to 3959)
AUTHORS Hasper,A.A., Dekkers,E. and de Graaff,L.H.
TITLE Direct Submission
JOURNAL Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
6703HA, The Netherlands
FEATURES
source
1. 3959
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
<1038..>3841
/gene="eglC"
join(<1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..>3841)
/gene="eglC"
/product="endoglucanase C"
join(1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..3841)
/gene="eglC"
/codon_start=1
/product="endoglucanase C"
/protein_id="AAK77227.1"
/db_xref="GI:19879406"
/translation="MRRTPAALVAGYLLDSVHAASQAYTWKKNVTTGGGGFTPGIVF
NPSAKGVAVARTDGGAYRLNSDDTWTPLMDWANNNSNHDWIGIDAATDPVDTRVYV

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 3,27e-74 Length: 3959
Score: 1620.00 Matches: 352
Percent Similarity: 55.57% Conservative: 107
Best Local Similarity: 42.62% Mismatches: 255
Query Match: 40.14% Indels: 113
DB: 8 Gaps: 20

US-09-917-376-3 (1-740) x AY040839 (1-3959)

Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIle----- 13
Db 1095 GCCGATCGCAGGCATATACCTGGAGATGTGGTCAC-AGGGGGTATCACATATCTTCG 1153
Qy 14 -----GlyGlyGlyPheValAspGlyIleVal 23
Db 1154 AATATGAGGGATTGCAAGCTAATTGATCAGCGCGCGAGGCTTCACGCGCGGTATCGTC 1213
Qy 24 PheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyGlyMetTyArg 43
Db 1214 TTCAACCTTCAGCTAAAGGTGCGCATATGTCGCACCGACATCGCGGGTGCATATCGG 1273
Qy 44 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 63
Db 1274 TTGAAC---TCGGACGATACATGACCCCGTTGATGGACTGGCCCAACACTCTAACTGG 1330
Qy 64 -----GlyTyr-Asn-----GlyValVa 69
Db 1331 TACGTAAGCTTCTCTTACGAGATGCTAGGCTCACTGTCAGGACGATTTGGGCGATTCA 1390
Qy 69 lSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThr 89
Db 1391 TCGCATCGCAGACCCCGCGTACGATACCGATCGGGTATATGTTGAGTTGGCATGTACAC 1450
Qy 89 rAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTr 109
Db 1451 CAACGACTGGATCCCAATGACGGCTCCATCTCTGCGGTGACCGACCGAGCGCATACCTG 1510
Qy 109 pGlnIleThrProLeuProPheLysLeuGlyAsnMetProGlyArgGlyMetGlyG 129
Db 1511 GGAAGACAGCAAGCTTCCCTTCAAAGTTGGAGGAAATATGCCCGCTGGGGTGTGGCGA 1570
Qy 129 u-----ArgLeuAlaVala 134
Db 1571 GGTAAGTCTTATATCTGGGTTTGTGTAGCATACACTGACCGCTGCAGCGCTCTGGCTGTG 1630
Qy 134 spProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTrpArg 154
Db 1631 ATCCGATGACCAACAGCATCTGTACTTTGGTCCCGAGTGGCAATGGCTTGTGGAAA 1690
Qy 154 erThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTy 174
Db 1691 GCACGGACTACGGCGAGACATGGTCCAATGTGACTGCCTTCAAGTGGACCGCACCTATT 1750
Qy 174 leAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVala 194
Db 1751 TCCAAGATTCCAGCTCGACC-----TACAGCTCTGATCCCGTGGGAATTCGCTGGGTCA 1804

```

```

AVGMYTNDNDGDSILRSTDOGTWETKLPFKVCGNMPGRGVCGERLAVDPDNDNSIL
YFGARSGNLWKSIDYGETNSNTAFKWTGTYFDQSSSTYSDPVGIVGIVPDSSTGSG
SGSPTRIPFVGVDVTGESVFSVDSAGETWVSGEPMYFLPHKGLISPSERTLISY
SNAGPYDLNGTGVKHTNIGSVMTDISPTSMIDTYDYGGLAVLDQVPGTVMVAALN
CWPDELILMRSTDGSGTWSPIWANGYPSINYYSYDINSAPWLQDDTSTDFPVRVG
WVVEALADIPDSDHMLYGTGETLYGGHDLQWMDSEHNVTLIESLAVGLEMAVLGLIT
PPGFPALLSAGVDDGGFHTSLTAPSYHYHTYSSYNTNGDILMNSKANPANSVRSGSD
SDPTLALUSSFGESWYADYAASSSTATQVALSADADTILMDNSKANPANSVRSGSD
VSSLPAGVITASDKANNITYFGAGSSFYLSGDSYTAATFTVTTLGSSTTANARAQPS
LAGDVWYTDITGLFHSNTYKGSFTQIGSGCTEGWSFGKPSDGDYVLPVFGFTVDG
VTGLFKTEDQGVNQIILSDAHEGFGSASANUVNGDLQNGRVFVCTNGRGIIFYGPPSG
TLPSATATASASSTAVKSTSTSTSKVSGSTVSSSTATITITSISLTITLTTTKS
SSSTTSTSTATATASATYGCQGGSGFTGTQPCSGWTCTYENIYYISQCKSIPIATDR
G"

```





[illegible]

Db 1540 GGCAACACCGCGGCACGCAACAG-----GTGGCCATCTCGTCGACGCGC 1584  
Qy 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla 531  
Db 1585 GGC CGCGCTGGACATCGACTACGCGCGCCACACGCTCCATGACGCGCGCACGGTGGCC 1644  
Qy 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551  
Db 1645 TATTCCGCGCGACGCGACGATCTCTGCTCGACGCTCGTCGCGC-----GTG 1695  
Qy 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571  
Db 1696 CAGCGCTCGCAGTTCAGGCGCGCTTTGCGCTCGAGCTCGCGCGCGCGCGCTC 1755  
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591  
Db 1756 ATCCCTCGGACAGACCAACACAGGCTTTTACCGCGGCTCCGATCGACCTTTTAC 1815  
Qy 592 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611  
Db 1816 GTCAGCAAGGACACCGCGCAGCAGCTTC-----ACGCGCGG---CCCAAGCTGGC 1863  
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629  
Db 1864 AGCGCAGGACGATCCGCGGATATCGCTCTCACCGACCCACCGCGGCGACGTTGTATGTC 1923  
Qy 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648  
Db 1924 TCGACGACGTGCGCATATTCGCTCCACAGACTCGGCGCAGACCTTTGGCCAAAGTCTCC 1983  
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlyLysSerSer 668  
Db 1984 ACCGCTGACCAACACTACCATCGCTCCGCTGCGGCTGCGGCTCA---GGCTCGAATCGG 2040  
Qy 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686  
Db 2041 ---AACGTGTATGCTTCGCGCACC-----GGCCGCTCAGGGGTGCGCTTACGCCAGT 2091  
Qy 687 AspAspCysGlyThrThrTrpValLeuLeuAsnAspAspGlnHisGlnTyrGlyAsnTrp 706  
Db 2092 GGAGACACGCGCGCTCTCTGACGACATCCAGGCTCCCGGCGCTTCGCTCCATCGAC 2151  
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly 726  
Db 2152 AGCACCAGGTGCGCGCAGCGCAGCAGCAGCGCGGCAAGTCTACGCGGCGCACCAACGCG 2211  
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly 740  
Db 2212 CGGGCGCTCTTTACGCTCAGGGAACGCTCGCGCGCGCGCGCGG 2259  
RESULT 12  
AKI10506  
LOCUS AKI10506 2724 bp mRNA linear PLN 24-JUL-2003  
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:002-167-F04, full insert sequence.  
ACCESSION AKI10506  
VERSION AKI10506.1 GI:32995715  
KEYWORDS FLI\_CDNA; oligo capping.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1  
AUTHORS The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Mizuno, K., Narioka, R., Aizawa, K., Hara, A., Hashizume, W., Masuda, H., Miura, J., Kishimoto, N., Nishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Ohtsuki, K., Li, C., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,

Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., KIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Itoh, M., Kogawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
Science 301 (5631), 376-379 (2003)  
2752273  
12869764  
2 (bases 1 to 2724)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Ikeda, R., Imamura, K., Imoto, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narioka, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Ooka, H., Osato, N., Ota, Y., Ohtsuki, K., Ooka, H., Ooka, H., Ooka, H., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shishiki, T., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.  
Direct Submission  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : http://cdna01.dna.affrc.go.jp/cDNA/  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.  
FAIS Genome Sequencing & Analysis Group: Ohtsuki, K., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narioka, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kaga, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sasaki, D., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sugabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.  
Location/Qualifiers  
1. 2724  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"



QY	652	SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerTyrProAlaVal	671
Db	2035	AGCGACACGTACCAGATCCGGTTGGCTCGCTCT---GGTTCCAACCTGG---AACCTG	2088
QY	672	PheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThr	691
Db	2089	TACGCCTTTGGCACGGCGCGCAACCGCGCCAGCGCTGTACCGCAGCGGACGCCGGCTCG	2148
QY	692	ThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGly-----AsnTrpGlyGlnAla	709
Db	2149	TCCTGGACTGATCTCAGGGCGACAGCGGGCTTCGGCATCATCAAC---GGATGCAGG	2205
QY	710	IleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIle	729
Db	2206	CTCGCTGGCAGTGGGAACGTCGCGCAACAGGTATACGTCGGGACCAACGCCCGCGCGCTC	2265
QY	730	ValTyrGlyAsp-----IleGlyGlyAlaProSerGly	740
Db	2266	TTCAGCGCGACGGTTCCATCTCCGAGCGCGCGCAGCGCCGCTCGT	2316
RESULT 13			
AP005031/c			
LOCUS	AP005031	299550 bp DNA linear BCT 10-MAY-2003	
DEFINITION	Streptomyces avermitilis genomic DNA, complete genome, section 11/30.		
ACCESSION	AP005031	BA000030	
VERSION	AP005031.1	GI:29606108	
KEYWORDS	Streptomyces avermitilis MA-4680		
SOURCE	Streptomyces avermitilis MA-4680		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1		
AUTHORS	Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.		
TITLE	Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)		
MEDLINE	21477403		
PUBMED	11572948		
REFERENCE	2		
AUTHORS	Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.		
TITLE	Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis		
JOURNAL	Nat. Biotechnol. 21 (5), 526-531 (2003)		
MEDLINE	22608306		
PUBMED	12692562		
REFERENCE	3 (bases 1 to 299550)		
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kushida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nite.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)		
COMMENT	This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Norihiro Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).		
	Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.		
	*1 Kitasato Institute for Life Sciences, Kitasato University		
	*2 National Institute of Infectious Diseases		
	*3 The kitasato Institute		

\*4 National Institute of Technology and Evaluation  
 \*5 School of Science, Kitasato University  
 \*6 Institute of Medical Science, University of Tokyo  
 \*7 RIKEN, Genomic Sciences Center  
 Following url is also available.  
 http://avermittilis.ls.kitasato-u.ac.jp.

FEATURES  
 source  
 1. .299550  
 /organism="Streptomyces avermitilis MA-4680"  
 /mol\_type="genomic DNA"  
 /strain="MA-4680"  
 /db\_xref="taxon:227882"  
 /note="This strain is also named as strain: ATCC 31267,  
 NCIM 12804 or NRRL 8165.-synonym: Streptomyces  
 avermectinius"  
 complement(42. .2936)  
 /gene="nrd"  
 complement(42. .2936)  
 /note="SAV2461"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative vitamin B12-dependent ribonucleotide  
 reductase"  
 /protein\_id="BAC70172.1"  
 /db\_xref="GI:29606109"  
 /translation="MTETASGPARSGRAKTKAKGLRIERTHTTGPVHPDEVWEVARR  
 DVVYTNWRDGSVNFGEORGVEFFDFWSNVAVNIIVTSKYGAVGTTPQREVSLROLIDRI  
 VYTKRAGDEYKIFASPADAEIIFEHLAYALLHQIFSFNSPVNFNWTGPQOQVSACF  
 ILIADDSMESILDWYKEGMIFFGGSGAGLNLRSKRELLSSGNSAGSPVSGFMGAG  
 DSAGTIKSGGATKRAAKWVILDVDPHDIEDFIETKYKEEKIIRALRDAGDFMDLGD  
 DITSVOYQANNSVRVNDTFMKAVEEGKFGLTSRMTGEVIEEDVAKSLFRKMAEAA  
 ACADPGIYDDTINWHHTCPESGRINSYEMHLDNTSCNLASILNLMFKLKDCK  
 GRQSFEVPSFAKTVLITAMDISICFADPPTQIKENTWAPROLGIYANLGALEMA  
 TGHVSDSGRALLAGAITSLMTGTSYKRSNELAAVPGPDGYARNAPQPHORVMKQESD  
 ANGVAVRDLDTPIWAAATEAWQDVHLHGKNGFRNAQSVIAPTGTIGLAWSCDTT  
 GLEPDLALVFKFLKVGGSQMIVNGTVPQALRELQVGEQIEAIVAHIAENGNIIDRI  
 GLKHEHYEYDFCAMGERSISAMGHVRMAAIQPMISGALSKTNLPETATVEDVEEYV  
 FRAWKGVKALAIYRDNCKVQGPLSAKKETAEVTAKTTEATIRAEKVEKVPYRVR  
 KRLPKGRPGITTSFTVGGREGYMTANSYDPDGIQVGEFLKMSKOGSTLAGMMDAFAIAV  
 SVLQGVPLETITVSKTNNRTEPAGMTDDPDVVRMAQSIVDYIFRLALDLFLPEFRS  
 ALGIHSAERORHLEGTSGPESDDDVDEGLAQSAPRAQELKAVATFPAEVAAPVPA  
 PQAKTASBELVENQLGIQADAPLCFCSGTKMQRAGSCYICEGCGSTSGCS"  
 complement(3086. .3649)  
 /note="SAV2462"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative regulatory protein"  
 /protein\_id="BAC70173.1"  
 /db\_xref="GI:29606110"  
 /translation="NHCPFCRHPDSRVDSRTTDDGTSIRRRQCPDCSRRTTVEFC  
 SLMVYRSVGTTEPFSFTKINGVRKACGPPTEDALAQIQGVEEAVRATGSAELTT  
 HDVGLALGLRLDLVAYLRASVYTRPQVLESDLEFMAIARLREERQRRPFAVDDHED  
 DAGAEQRQDTRGSGGTVEVPVATVAD"  
 4168. .4947  
 /gene="lexA"  
 4168. .4947  
 /note="SAV2463"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative SOS regulatory protein LexA"  
 /protein\_id="BAC70174.1"  
 /db\_xref="GI:29606111"  
 /translation="MTTIDASATITAQDSRQGLRLEPVHMANEATSHGPKRSLPGRRP  
 GIRADSGTIDRQRVIEIVRDSVQRGYPFSMRLEIQAVGLSSTSVAHQLWALEKRE  
 GFLRDPHPRPAYEVRGQSQSSVQPTDAGKPAASYPLVIGRIAGGPTLIAEESVD  
 VFLRPLQVGDGLFLVKVYGVNEMIAAICDGDWTVRRQPVVAENGDIVAAMLDEGAT  
 VKRFKREDGHWLPLPNSAYQRIPEGDEATILGKVVAVLRV"  
 complement(5110. .7143)  
 /gene="ding"  
 complement(5110. .7143)  
 /gene="ding"

gene  
 CDS

gene  
 CDS

gene  
 CDS

\*4 National Institute of Technology and Evaluation  
 \*5 School of Science, Kitasato University  
 \*6 Institute of Medical Science, University of Tokyo  
 \*7 RIKEN, Genomic Sciences Center

Following url is also available.  
<http://avermittilis.is.kitasato-u.ac.jp>.

## FEATURES

gene

CDS

CDS

gene

CDS

gene

CDS





Db 160613 GCCACGGGGAGTGGACCGAGGTACCCCGGTGAGCGCGGGGAGCAGCAGCGACGGC 160554

Qy 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313

Db 160553 TCGCGGACACCTTCGCTACGGGGGTGCGCGTGCAGCGCCCGCGGACCCCTC 160494

Qy 314 MetValAlaThrGlnIleSerTyrTrpProAspThrIleIlePheArgSerThrAspGly 333

Db 160493 GTCGTCTCCACCAACACCGCTGGCGGCGAGCGGCACACGGGTCTTCCTCCGTCCACGACGGC 160434

Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353

Db 160433 GCGGTACCTGGAGCTCCCTCAAGGAC-----GCCGCC 160401

Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373

Db 160400 GTGTTGAGGTGTCGAGACTCCCTTCCTGACTGGGGCGAGAC----- 160356

Qy 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393

Db 160355 ---AAGCGAGTTCGGCTGGTGGATCCAGCGCTCGCGGTGCGACCTACGACTCCAG 160299

Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413

Db 160298 CACGTGCTGACGGGACCGCGGCGACCTCTACGGCACCCCGGACCTCAAGCGCTGG--- 160242

Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433

Db 160241 -----GCACCGCGATCCGCGCGCTGGAGGAGAGCGCGTGGCG 160203

Qy 434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452

Db 160202 CAACTGATCTCGCCCGCTCGGGAGGACACCTGATCAGCGACTCGGGGACATCGGT 160143

Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471

Db 160142 GTGATGTACACAGGAGCGCTCACGGCGTCTCGTGGCGGCGCATGGGACGACCCCGTG 160083

Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491

Db 160082 TTCGGTTCGCGGACGAGACTCGCGAGCGCGCGCCAGCGGGGTATGTCGTCGCGACG 160023

Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511

Db 160022 GCTCGGCGGACACCGGACCGC-----GGTACTCCACGACGCGC 159981

Qy 512 GlyLysAnTrp-----PheGlnGlySerGluProGlyValThrThrGlyGlyThr 529

Db 159980 GGGCGGACCTGGGCGCCTTCGAGGCCCGCCGACATCGCCAGGACGACCGGACCG 159921

Qy 530 ValAlaIleSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545

Db 159920 ATCGCCACAGTCGCGGCGGACACATGCTGTGTGCTCTTCGTCGACCTGGGACGCGC 159864

Qy 546 ProGlyGlnProValValThrAla-----ValGlyPheGlyAsnSerTrpAla 561

Db 159863 -----ACGAGTACCGCGCCCGCCCGTCTCGAGCGGACACCGCGGAGTGGTCC 159816

Qy 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581

Db 159815 GAGGTCTCTCTCCGAGGCGGCGCACCGGTGCGCGACCGCGGCGATCCGACGCGC 159756

Qy 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599

Db 159755 TTCTACGCTACGACTTCGACAAATGAACGCTATACGCCGACTGACGATGGCGGTTCG 159696

Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619

Db 159695 TTCACGCGCGGTGGCGGAGACTCGCC---TCCGGCGACACGCCAGTTCAGCTGGTCCG 159639

Qy 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638

Db 159638 GCGCGGAGCAGACGCGGACCTGTGTGCTCTCCGCCAAATGGAACGGGCTCTACGCTCC 159579

Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658

Db 159578 ACCGACGGCGGGACACCTTCGCCAGGATCGACAGCTGTGGCGCTCGTACACCCCTCGC 159519

Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678

Db 159518 TTCGCAAGGCGCGCGACCGGCGGACCTACCGCGGATCTACAGTCTCGCTCGACGGAG 159459

Qy 679 GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAsp 698

Db 159458 ACCATCACCCTGCTACGCTCCGACGACGCGCGCGGACGATGGTCCGGATCAACGAC 159399

Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718

Db 159398 GACGCCACCATGTCGGGTGGATCGCGAGCGCGTCTCGTGTGACCGCGCATCCACGGC 159339

Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733

Db 159338 CGGTCTACCTCGCCACCAACGCGGCGCGCATCCAGTACGGGAG 159294

## RESULT 14

AE011809 14520 bp DNA linear BCT 29-MAY-2002  
 LOCUS Xanthomonas axonopodis pv. citri str. 306, section 187 of 469 of  
 DEFINITION the complete genome.

ACCESSION AE011809 AE008923

VERSION AE011809.1 GI:21107967

## KEYWORDS

SOURCE Xanthomonas axonopodis pv. citri str. 306

## ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Xanthomonadaceae; Xanthomonas.

## REFERENCE

1 (bases 1 to 14520)  
 da Silva, A.C.R., Ferro, J.A., Reinach, P.C., Farah, C.S., Furlan, L.R.,  
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
 Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
 Chamergo, P., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,  
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R.H.A.,  
 Rassi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
 Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with

differing host specificities

Nature 417 (6887), 459-463 (2002)

JOURNAL MEDLINE 22022145

PUBMED 12024217

## REFERENCE

2 (bases 1 to 14520)

da Silva, A.C.R., Ferro, J.A., Reinach, P.C., Farah, C.S., Furlan, L.R.,  
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
 Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
 Chamergo, P., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,  
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R.H.A.,  
 Rassi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
 Kitajima, J.P.

## TITLE

Direct Submission



JOURNAL Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

FEATURES source

1. .14520  
/organism="Xanthomonas axonopodis pv. citri str. 306"  
/mol\_type="genomic DNA"  
/strain="306"  
/db\_xref="taxon:190486"  
/notes="pathovar: citri"  
87. .2342  
/gene="cela"  
/notes="synonym: XAC1770"  
87. .2342

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="cellulase"  
/protein\_id="AAM36634.1"  
/db\_xref="GI:21107968"

/translation="MTAVIDPPTKRRBSAAHMLTAVCLTLVLLFWTSEASDERGTP  
DPQWMSVAGCGGFTVGNLFHPAERDLYARTDVGAYRWDAQAQWALTWLGLAD  
DNLWGLDPAVDPAFPNALYLAAGTIVHERAGNAVLRSDFRGKQFERADLPFLGG  
NQLGRANGERLAYDPDHGRVLLGSADGLWRSDDRGAHWARVEGFPADLAGATARN  
HVSGEQAAGTAFVVFDAAGHAGAATPRIYGVSTQATSLYVSEDAQTWSAVAGOPK  
GLRPSHMVGQOWLYSGDRPGDPLMAGGALWNYDQGRWRETSPIPOPATGDF  
GMCAPVDHPQOVLLASPRETDRPEVRSFGGSGRSWVLLAAQFQHPAPWTAH  
ATPHWIGALIDPFDNSHATPVYGYGIWASRNLTQTSQQPLOWFQDRGLEETVPLD  
LUSPMAGHLLSALGIDGRHDAALTAQYLGPRLTNGESIDGAGAPQWVVRSGT  
VDRNRNETALYSODGTHFAAFSEPRGQAGTIAAADAQVWVVDQGVWRT  
GDFKRWQVGLPDTAVVADVADQWVADRVGRGLTYESDSDGAEFDTGQOVS  
PARDEARQPLDPWRAGVYLAASPTLGVWQNGQLRTLKSDARSIGIKALRA  
GAPALYLAGRVAGVDFRSDDEGAHWRINDAHRFGKPSYVTGDPRIAGRVYFAT  
GGRGIFGDRP"

2431. .4347  
/gene="XAC1771"  
2431. .4347  
/gene="XAC1771"

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/transl\_table=11  
/product="sialic acid-specific 9-O-acetyltransferase"  
/protein\_id="AAM36635.1"  
/db\_xref="GI:21107969"

/translation="MTASFVRRGCLGLVLVSPCANAVPTLPLLLADGAVLQRDQMP  
VWGSSPNAIAVSPGKRATKADATGQWKLRLPAHAAGPYLVRVQGGDELQVRD  
VLVDVWLAGGQNMWELQAQSDGPOAVAAANDAOLRQPKPKWSVQPARLTGGE  
WKAATPANAGEFTAVGVFFAKELRSTGVPDGTVNSTWGGSAIEAMWDAASLGNADO  
NQGAIEALKORDAAQAATGKRLARWPKVEDENPQWREAFDDSDWDSIPVTKWESS  
GYDMDGIAWRTITLSAAEAAGGTLGVGQIDSDTITTYNQAGQVSGTEKQWNLPRV  
YQVPAALAKGVANHIAVRDELSGGMGHPDEQRFVTSAGAKRQDGLWKKFPAARV  
VSLTDKNQOLPVLVYNHLPLOFPFVKGVWYQGETNADTDGAVRYEQFAAMIRGW  
RABRDGKTLPLWLLKAGKAGDQKGLSPWALLRSQSKTLALPQFOQAVIIDIGNP  
TDHPNKRQDVGHRLAARHAYGETLVSAFVEKASPDGKAVLGFEDLQGSALQV  
RGGAVQGRFAGACDHPFATQIDGDRIVRSDAVAAPVARYGSENPPDANLIN  
RDALPVSPPFTFW"

4344. .6101  
/gene="XAC1772"  
4344. .6101  
/gene="XAC1772"

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM36636.1"  
/db\_xref="GI:21107970"

/translation="MTRSLMSMHSVPARVIASSPARRRRLPLLSALALAAACQBSGE  
QDDKAPAAVATAGSAATTPATAAPAAQTMPQFVSKDGKHALWVGAPFLILGAQ  
VNNSNYPGIMDKVPMQVLPNTVQVPIAWEQVBEPEGKDFDFVDTLLNQAREHK

gene

CDS

VLRLVLFATWKNNGPOYADPHWVXTNQRPVVTRECKSIGSLSPHAQATLADARKA  
FVAFWQGLKKVDPQHTVIMIQPENEAGTIGSVDRFSPWAQKLFQCVDELIGRWKQ  
PGTWAQGLADADEIFHWSIGRYIDQIAQAKAEYPLPMYVNAALRDFPFGPQGY  
ASGPTDNLVDVWKAAGHIDLLADPYIMPEYRMVTVLERYARPDNALFVAETGNRP  
EYARLYPTLGHNGIGSAFGIDTRYSNYPLGAKHVNVEETLAPFALGFAVSMGWRP  
PAKASGKLGHTAEESQVAVQELKNDWSATITYGVPQFWEKCEPGNPEPIGAAL  
IAELSPDEPLVTGHARVTLHPASDATANMIYDRVSEGVYEGNDWKFORWNWQDQTDY  
GVNFSVPQLLKIKLATYK"

5999. .9109  
/gene="xyls"  
/notes="synonym: XAC1773"

5999. .9109  
/gene="xyls"

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="alpha-xyloridase"  
/protein\_id="AAM36637.1"  
/db\_xref="GI:21107971"

/translation="WEIPAQLQEMRPDLRRQFFORAAAQQAQAGNLQISAPATTPRRR  
RDGIASVPYCAAPAAAPRHSKGVIMOTTRISAGAMPANRRTSLAQCLALSALLANA  
AHAQEVKKAQGVTVPSAKGAAVRLQVVDAGIIRVSADPDGDFARSPSLMRVPQG  
DTAQVAQSGDSVOLKTKGVKTASISTVDGHVSPADANGKPYLSEVAGSRFAPLNVG  
KQYLSVRQFQSPDEALYFGQHQQGMNOKRNVLCQNNIDNAPVLSRVNGL  
LWNNSTIRLGDPRGLQPLPKTLTYDAKAGALTAARYAINGKHLERRESEVYQY  
LSDITPKKAITKDSRMQVTEGTEIATLGGHTFSLYSSEYAKLVYDGLKVVDR  
WRONPNWHEFKLDLBPGRHTVKVWDLLDPSYALLHRDPLPAEAKDLSLWSEA  
QOMIDYFVSADSVDQAVAGYALTKGSTMPLKWAYGFWQSRERVKSDDELVGAEEY  
RKRKSLDNLVLSYMPENAWGSHDFDPQHPFDDGMVKAHVDMHAQIMISWPKFY  
PTTAYKELDAAGFMFRNEVEGELMIGKYGKNSFYDYPSEKAQAIYWRQINEKLS  
KGFDAWMDADEPDVHNLDAERKARTTNALGSSTFYNSYPLPHTHGVVYGRRAA  
DDKVFILSRKGYAGTORNNAVWSDGTVSRWDMRDQISGVNNAISGLPNMTDIFG  
GFAVEKEDQPAHLPEWRELNRWFQFAGVPIFRSHGQFPYREIWNIAPEGTPFY  
ESMAVYRLSALPYLSLADTYQDRGVTMRGMMDPNDPKYRDINDOYLFGPAF  
LVAVTRGATSRQVLYPAGSWLEFATGKRYEGQCSIEAAPPIERMELFVAGSIVP  
TGPVQYVDQVADPLTVVVTGADQGFSLYEDDGKGYEKGEFESFPLVWNAQGE  
LSIGKREGSWTMOAKRTINRVFDGPRDDAGALAPKTDASTIQYDGKPVSLQKRIAS  
GKARRR"

9201. .11573  
/gene="XAC1774"

9201. .11573  
/gene="XAC1774"

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM36638.1"  
/db\_xref="GI:21107972"

/translation="MLTRRELCKATGAIVLGVAPAAQAQATQASKTTALPVVAAAEA  
LOLWYRFPANWEVALPVNGRLGAMWGGIAHERLOLNEDTLVAGGPVDSYSPDALA  
ALPOVRALITFAGRYAEAEKADAKLSRPLKMPYQPLGDLILLDFRADGIDSYRQL  
DLDTAVATITFRSGGAVHRRFVCAQACIIVRLSCDRPGGIDSVRGIDSPQGEVT  
AEPGLDFGRNGSAGIEGRFALRVLPQVSGKLSQVRDLRIDDAADLYVLLLSA  
ATSQVRDADVDGDPALTAARLKAALDFFALLRAHLADHQLRFRVAIDILGSSAV  
QLPTDERVORFAGNDPALAALHYOGRYLLICSSRPGTQANTQCIWNLDLQWPWES  
KLTININENWYSEANALHECVPLEAMFLDLAQTCHTARAIVDAPGWVWNND  
LWRQAGTIDGQWSLWPMGVLVQLQDLWRWDYGRDAYLSKVYLPFGKAEEFVATL  
MRDPTGAMVTNFSMSPENHPGAAVCAQSPMDQLLDLFAQCIAMSKLIGIDAQL  
AQOALAREQLPNRIRIGAQOLOQWQMDMQAPEINHRVSHVSLHALPSSQINLRD  
PELAAAARSLREIRGDNATGIGWRNLNWARLADHAYRILOLLIPERTYPNLFD  
AHPFQIDGNFGGTAGITEMLLQSGGVSFLLPALPKAWPRGSRGLRVRGASVDLE  
WEGRLQOARLHSDRGGRYQLSYAGQTLDELGAGRTQOVGLNNRLVTQ"

11570. .13063  
/gene="XAC1775"

11570. .13063  
/gene="XAC1775"

/notes="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"

/codon\_start=1  
/transl\_table=11  
/product="D-xylulokinase"



```
Qy 647 -----AlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 663
Db 2067 CAGTTCGCCCGCGTATCAAAATCAGATGAAGCGCGCTCGCTGGGCATCGGCAAGCGGCTA 2126
Qy 664 ProGlySerSerTyrrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683
Db 2127 CGTCCGCCGCGCGCGCGCGCGCTGATCTGCCGCGCGCGCGCGCGCGCGTGGATGGCATC 2186
Qy 684 TyrArgSerAspAspCysGlyThrThrTrrValLeuIleAsnAspAspGlnHisGlnTyr 703
Db 2187 TTCGCTCCGACGACGAGCGGCCCATTCGGCGCATCAACGACGATGCCATCGCTTC 2246
Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGly 723
Db 2247 GGCAAA---CGGTATACGTCGACCGCGGACCGCGGATTCGCGCGCGCTGACTTTGCC 2303
Qy 724 ThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 2304 ACCGGCGCGCGTGGCATCTTCTATGGCGAT 2333

RESULT 15
AE001712
LOCUS AE001712 13278 bp DNA linear BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 24 of 136 of the complete genome.
ACCESSION AE001712 AE000512
VERSION AE001712.1 GI:4980799
KEYWORDS
SOURCE Thermotoga maritima
ORGANISM Thermotoga maritima
REFERENCE 1 (bases 1 to 13278)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
Eisen,J.A., Fraser,C.M. et al.
EVIDENCE for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
PUBMED 10360571
REFERENCE 2 (bases 1 to 13278)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Direct Submission
TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 13278
Location/Qualifiers
/organism="Thermotoga maritima"
/mol_type="genomic DNA"
/db_xref="taxon:2336"
150..1985
/gene="TM0300"
150..1985
/gene="TM0300"
/notes="similar to GB:AE000657 percent identity: 48.20;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein, putative"
/protein_id="AAD35388.1"
/db_xref="GI:4980800"
/translations="MKRFLFVAVLFTFAAFQFLNVPREQTIVVPGPKTANPGDN
VIAGWKDPHRIQQLLEPLWMDPPVRGEIINALAETGIYNEDFTMIIVKLKGVYV
SDGVPFATDDVFTVEMCKYPEMARHMEFNEIDVKYKEDDITVVFELKKNTRPHT
```

gene

CDS

```
/gene="TM0301"
2050..3054
/gene="TM0301"
/notes="similar to GB:Pyro h percent identity: 59.88;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD35389.1"
/db_xref="GI:4980801"
/translations="MKFTKSYLIPRLIQYVYVMVIGITVFLIPRLSPVDRPIQLNT
RLRQNTPEAVIKMREWEELVGLKGTWEOYVNLWKRLKGDGFGSPWVYYPASVIS
LIRNSLPWTGLLTTLTISVIGLIVGAILGYKESFGSKLTGILMAINIPYVIF
ALTILLFCYVPIPFPLGGSGVQKISLSWSFILDVILKHTLPALSLILGFGAGL
GKILVNQREEDIVVYARMMLRNSKILFSTVLNAMDQVIGLSLGGIFGGMW
TEILTYPGIGTLAYNAVANDYNMIAISFISIIAVATALLILDLLYPLDIRKYR
"
3073..3942
/gene="TM0302"
3073..3942
/gene="TM0302"
/notes="similar to GB:Pyro h percent identity: 57.09;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD35390.1"
/db_xref="GI:4980802"
/translations="MKRILRLFFDMWEDLRFKAVYVLTALVLLSILSIFSYPDPYK
VYLTNPQPSLEHILDTGKQDIPWLLTFLSKNSILATIAAFSRIITIGMLSG
YLGTDRILMALTDTPMVPVPELLALLSLIKNVLSPILGILGIVFGWAGDARVI
RSMILSRERFVTSKSGRTFVEVGFVPIYLPVIGSGFSGMGAIGFEIVLA
ILGTRVEIPTLGSMSFMHMINFQAMLLGYWVWLTPITVAVFLFTALYTLISUSINEYL
DPTRMQRIGAVK"
3945..4931
/gene="TM0303"
3945..4931
/gene="TM0303"
/notes="similar to GB:U08186 PID:473270 GB:AE000512 percent
identity: 76.00; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD35391.1"
/db_xref="GI:4980803"
/translations="WEKTAENLRVYFIKSGDKTTSIRAVDNVSLHFFENEVYGIA
GESCGKSTLKVLFGEIKPPLKLVDSGEVFKLDEEYRINSMLDQLKKIKWYVS
VPGSGSVLNPVRKIRLFLDIVBEHYSMSREAFSIESFIEALGPLVRVDSYSHQ
LSGMQRORVATATVLPKVIADPTTALDVVQVGVIGLKKIKHKEQKTLVIT
HDMGVRHARLATRIGVYAGRLIEAEETHEIFKNPLHPTFRFLNSLPRGDRSRKEGT
PGSPPSLANLPSCGPFHPRCFADQICMEMEPDLIEVARNHKVACWKVGVER"
4928..5878
/gene="TM0304"
4928..5878
/gene="TM0304"
/notes="similar to GB:Pyro h percent identity: 65.51;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD35392.1"
```

```

/db_xref="GI:4980804"
/translation="MRVSVNRLTKIFRVGSIFFSLKVLADNVSVFSGSEITPSLAGS
SGCKTTAKILLGDSVEPTNVEHFLYDGIKLKDFVEMEHFLKKVQVTVQONPSTFNP
RYEDYLFKTVLFRKMAVNTDEASRLTIEKLSVGLNFQIRIKGRYSEFSGGQLOIRS
IARALLTNPSFLVADPEMSIDALRTISIVNLFRLDKERRGISILYITHDLSFAYVVA
DKLAIMRGSIVEMGQAQDILDTPPHPYTKLLKSSIPVDPDPEYEMDENINLSSTEREE
FFAKGCKFASRCPEVWSICRESVPPDPVNGRVVKCWLK"
5894..8017
/gene="TWO305"
CDS
5894..8017
/gene="TWO305"
/notes="similar to GP:356020 percent identity: 55.11;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase, putative"
/protein_id="A035393.1"
/db_xref="GI:4980805"
/translation="MLRSFLILAILGVVFGATFEKWSVEINGGVFPGIIFHPASP
GLLYARTDVGLRYRDEETKRWQLDFLRDQSDYMGVLSVALDSDPKRIYAMTKG
YTDWAGYGAILSEDTGWTWITNLIDKYGKVGNEDEGNACERLQVDPNFSVLGPF
GTYKGLKXSEDEGKNNKVDSPSSVTFLVDEKSGEGKSPTPIR FVGCSEPKGIF
VTEGGTWNVLNPLNDLILPGRKIDHGLIYVTLNLAGPNGATRGAVNKKYVIADQK
WYDTPMKGDGYCYQDQNVNVIATLDRWYDHDVETFLSDGETWRPLLEKANFDI
NKAPMTKDLNPHGISVDKVIDPDMNRAIFTTGYVMVTVELKKSPFGMGKPKVKMIFEN
RGLSEVHLTVLPPIIGERPLLISAIDAGGRFLESLOTDPSSMYKPLWTKSLGIAFAYQ
NSKFARVHTYVTFLLSYSDSGGINNREIETVPEGITDGRLLSVNDGKTLAWGSA
NHEVIVSSDKGSKWKALSYVPEFNYPRASDPVNSKXYLPDMKNGDPLISKDGKKS
PMKAGLPSPDNWVLSYFPLVADPREGDIWLALQNGLYRSKDGKITFERLGNVDI
AYVIGKAPPGGTDTYPAIYLGNMVGWYGVGIFMTSGDEGTWMRINNDKHQHFTHYIMG
DMNEFRGIFLGTGREGIIVGEVKEE"
8014..9363
/gene="TWO306"
CDS
8014..9363
/gene="TWO306"
/notes="similar to GB:M29877 SP:P04066 PID:1335066
PID:178409 PID:182779 percent identity: 56.91; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-L-fucosidase, putative"
/protein_id="A035394.1"
/db_xref="GI:4980806"
/translations="MISMKPRYKPDWESLREHTVPKFDKAKPGIFIHGVIYSPGVA
TPTEGLGKPMDAWFCQNPVAYEVNSLRIKESPTWEYHVTKYGENPEFQADLFTA
EKWDQENWADLFKAGAYVITPTKHHDFCLGTYKTPDPSVKRGPKRDLVDLAKA
VRAGLRFGVYISGGDWRTEPTPIRPEDLSYRNTPEYADYATYKQWELVDLYLP
DVLNMDMGPEKEDLKYLFAYVYKHPGSGVNDVRGPHWDKFTAAHYVNPGLD
KYWEYETRGILGFGYRNNGEPHMLISVEQLVTVLVDVWSKGNLLANVPKDGDTPI
DOKERLILGELWRLYKGDAIYGTYSVWERCAKTEGTELRPTRKCNRI FVIFLGITPT
GEKIVIEDNLISAGTVRHVFTGERLSFPKNVKGKLEITVPKULLETSIILVLEAVEE"
9360..10781
/gene="TWO307"
CDS
9360..10781
/gene="TWO307"
/notes="similar to GB:L42023 SP:P44779 PID:1005445
PID:1220691 PID:1204865 percent identity: 56.91;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-fucose isomerase, putative"
/protein_id="A035395.1"
/db_xref="GI:4980807"
/translation="MKGYKVGILSPSDGREYVHRELEGIKFKPEDEIAKTIETGEV
EVIIRAKEI VWKPSIAKREARLAEAGEVTFINYA WAFPHFTVLASKAPGFLFS
NINPQYGMVMSIAAGALQDQGTGKRVMMGNTKDEKLVKVMVAFTRAGTGYKLLRG
RYGVGSGMGTAVPNVDLWNKI FGVDVHIDQFEI VRRSOLI PDERARKGLMIE
EKAVAHYDGKYLTPKLELQIKSYHAVREI VEEMELDFVGI KQOLELTFHFVTMVDI
ETFLNDPYWEGSHEPIVCATEADSOGLTQMQIFLKIATPVLFAVRHYVEZYDVL
LCNSGNHATYFAARSFDDPNMCKKVFYPTQIFYPAGGAARVHRIAPAGKVTYSLRLTR
DGRYRTVPVFGFDEPEENKKNYIADSIQNNWPHAFLEKMETPIDEFALKVYSSNHIGH
YGDYVEIKTFCBIASIDFVLMK"
complement(10798..13092)

```

Db 6725 TAGCATGTAACCAATGAAGAGAT-----TTCGGATATTGTGA 6766  
 Qy 303 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 322  
 Db 6767 ATAGATGTTCAAGAG-----AACGTGGTATAGTATCCACACTCGATAGTGTAT 6817  
 Qy 323 ProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAsp 342  
 Db 6818 CCACACGATGAGATATTATTTCCTCGACGGAGGAGAAACGTGGAGACCTCTCTCTGGAA 6877  
 Qy 343 TrpThrSerTrpProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 362  
 Db 6878 AAGCGAAATTT-----GATATCACAAAGGCCCTCG 6910  
 Qy 363 LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp 382  
 Db 6911 ATAANA-----GATTTAATCT-----CACTGGATATCT 6940  
 Qy 383 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyThrGlyAlaThr 402  
 Db 6941 GAC--GTCAAGATAGATCCATTGACATGAATAGGGCTATTTTCAACACAGGTATGGT 6997  
 Qy 403 LeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaPro 422  
 Db 6998 GTCTGGGTGACTTATGAGTTAAAAAAGTCATTCGAAGT-----ATCGGAAAA 7045  
 Qy 423 MetValLys-----GlyLeuGluGluThrAlaValAsnAspLeuIle 436  
 Db 7046 CCGTGAATGGATATTTCGAAATCGAGGGTTAGAGAAACAGTTGTTTGCACCTGGT 7105  
 Qy 437 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456  
 Db 7106 CCCCCATAGACAGAAAGCCCTCTCTCTGCAATCGCGATGGGGAGGATTCGGTCAC 7165  
 Qy 457 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 476  
 Db 7166 GAAAGTTTGGACACCCACCATCTATCTATGAC---AAACCCCTAAATGGACGCTCTCT 7222  
 Qy 477 SerValAspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 496  
 Db 7223 GGTATAGCTTTTGCCTATCAGAAATCAAAATTTGGCAAGAGTGCAATACATAC--- 7276  
 Qy 497 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe 516  
 Db 7277 ---ACCTATCT-----TTCTTGCTTACTCCGAGGATGGCGGAATCAACTGGAGA 7324  
 Qy 517 Gln---GlySerGluProGlyGlyValThrThrGlyGly-----ThrValAlaAlaSer 533  
 Db 7325 GAGATAGAAACTGTCCCGGAGGGAATAACTGACGGCGGAGGTTAAGTTTGGCAGTCTCG 7384  
 Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyAla 553  
 Db 7385 AACGATGGAAAAACACTGGTCTGGAGTCCAGCGAAT-----CATGAGGTAATTTGTTCA 7438  
 Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla-----AsnAla 570  
 Db 7439 AGCGATAAGGAAAAAGCTCGAAAAAGCCATTAGTGTCCAGTCCAGATTCAACTAT 7498  
 Qy 571 GlnIleArgSerArgValAsnProLysThrPheTyAlaLeu-----SerAsnGly 588  
 Db 7499 TTTCTCTGCATCTGATCCTGTAATCCGAGTAAATTTTACATCTTCGATTGGAAAAACGGT 7558  
 Qy 589 ThrPheTyArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608  
 Db 7559 GACTTCTCTATAAGTAAGATGGAGAAAGAGTTTATGAAA---GGTGCTAAACTTCCT 7615  
 Qy 609 SerSerGlyAlaValGlyValMetPheHisAla-----ValProGlyLysGlu 624  
 Db 7616 TCTTTGATAACTGGTGGGTATCGCTTTATCTTCTCTGTTGGTCTCTGACAGAGAA 7675  
 Qy 625 GlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyThrHisThrAsnGlyGlySer 643  
 Db 7676 GAGATATCTGGTTAGCACTTCAATGGAATGGACTTTTACAGATCAAAAGATGGTGAATA 7735

Qy 644 SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 663  
 Db 7736 ACGTTTCAGCCGACTAGGAAACGTTTGATATAGCTACGTGATGGTTTCGGAGCCCCAAAA 7795  
 Qy 664 ProGlySerSerTyProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683  
 Db 7796 CCAGGTACTGACTATCCAGCAATTTTACCTGAACCGGTATGGTGAATGGAGTTTATGTTAT 7855  
 Qy 684 TyArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTy 703  
 Db 7856 TTCATGTCGACAGATGAGGCAACACATGGATGAGAATTACACGATAAGCATCAATT 7915  
 Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyThrIleGly 723  
 Db 7916 GGA---TGGATACATTACATGATAGGAGATATGAATGAATTCGGGAGAAATTTTCTTGGT 7972  
 Qy 724 ThrAsnGlyArgGlyIleValTyGlyAspIle 734  
 Db 7973 ACGAAGGAAGAGGAATCATCTGTTGGTGAAGTG 8005

Search completed: May 12, 2004, 09:25:37  
 Job time : 11278.8 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 1026.48 Seconds

(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIVGDIGGAPSG 740

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124095041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/US09-917-376-3/runat\_11052004\_114534\_28231/app\_query.fasta\_1.11429  
-DB=N Geneseq 29Jan04 -QFWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN\_1\_1758 @runat\_11052004\_114534\_28231 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	2869	7 ABZ77632	Abz77632 Nucleotid
2	4036	100.0	2869	9 ADD22922	Add22922 Acidothe
3	2063	51.1	2823	7 ABV76941	Abv76941 Nucleotid
4	1506	37.3	2849	9 ADD42060	Add42060 Trichoder
5	1139	28.2	3668	9 ADD24893	Add24893 DNA encod
6	878.5	21.8	5698	9 ADD42054	Add42054 Trichoder
7	392.5	9.7	1103	3 AAF14988	Aaf14988 Trichoder
8	238	5.9	7407	7 ACA42281	Aca42281 Prokaryot

9	238	5.9	7407	8 ACC59398	Acc59398 Microbial
10	199.5	4.9	7305	7 ACA23399	Aca23399 Prokaryot
c 11	195	4.8	5295	7 ACA24001	Aca24001 Prokaryot
c 12	192	4.8	9903	7 ACA40315	Aca40315 Prokaryot
c 13	192	4.8	110000	4 AAI99682_04	Continuation (5 of
c 14	184	4.6	3324	7 ACA38418	Aca38418 Prokaryot
c 15	184	4.6	110000	4 AAI99683_04	Continuation (5 of
c 16	182	4.5	11148	7 ACF39394	Acf39394 Mycobacte
c 17	182	4.5	110000	4 AAI99682_37	Continuation (38 o
c 18	182	4.5	110000	4 AAI99683_37	Continuation (38 o
c 19	181.5	4.5	11679	8 ADA30090	Ada30090 DNA encod
c 20	179.5	4.4	3222	7 ACA26562	Aca26562 Prokaryot
c 21	177.5	4.4	110000	4 AAI99682_30	Continuation (31 o
c 22	177.5	4.4	110000	4 AAI99683_30	Continuation (31 o
c 23	176	4.4	32367	2 AAV35620	Aav35620 Human SHO
c 24	174.5	4.3	67459	7 ABQ77491	Abq77491 S. aurant
c 25	174	4.3	2855	4 ABL29005	Abi29005 Drosophil
c 26	174	4.3	5221	4 ABL19628	Abi19628 Drosophil
c 27	174	4.3	5302	4 ABL29004	Abi29004 Drosophil
c 28	173	4.3	2708	4 ABL19629	Abi19629 Drosophil
c 29	170.5	4.2	2176	4 ABL11445	Abi11445 Drosophil
c 30	170.5	4.2	4252	4 ABL11444	Abi11444 Drosophil
c 31	170.5	4.2	7191	7 ACA47178	Aca47178 Prokaryot
c 32	170.5	4.2	25165	8 ACD19026	Acad19026 E. coli
c 33	170	4.2	4323	7 ACA20513	Aca20513 Prokaryot
c 34	170	4.2	110000	4 AAI99683_37	Continuation (38 o
c 35	169.5	4.2	3360	9 AAD62307	Aad62307 Candida a
c 36	169	4.2	110000	4 AAI99682_37	Continuation (38 o
c 37	169	4.2	110000	4 AAI99682_38	Continuation (39 o
c 38	168.5	4.2	25473	9 ADC01011	Adc01011 Enteroha
c 39	168	4.2	5451	6 ABZ35486	Abz35486 Human gen
c 40	166	4.1	3879	7 ACA20847	Aca20847 Prokaryot
c 41	165.5	4.1	4383	9 AAD62312	Aad62312 Candida a
c 42	165	4.1	6615	7 ACA40310	Aca40310 Prokaryot
c 43	165	4.1	110000	4 AAI99682_03	Continuation (4 of
c 44	165	4.1	110000	4 AAI99683_03	Continuation (4 of
c 45	164.5	4.1	3237	7 ABT42169	Abt42169 Toxicity

#### ALIGNMENTS

RESULT 1

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX AC ABZ77632;

XX DT (first entry)

XX DE Nucleotide sequence of the avicelase AvIII.

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulose; biofuel;

XX KW detergent; pulp processing; paper processing; feed processing; textile;

XX KW cellulose; gene; ss.

XX OS Acidothermus cellulolyticus.

XX PH Key Location/Qualifiers

XX FT CDS 1..2869

XX FT /\*tag= a

XX FT /partial

XX FT /product= "AvIII"

XX FT /transl\_except= (pos:2869,aa:Xaa)

XX FT /note= "Xaa is an unspecified residue"

XX PN WO2003012090-A2.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023818.

XX PR 28-JUL-2001; 2001WO-US023818.

(MIDE ) MIDWEST RES INST.

Ding S, Adney WS, Vinzant TB, Himmel ME;

WPI; 2003-248177/24.

P-PSDB; ABP73015.

New thermostable AviiiI peptide from *Acidothermus cellulolyticus*, useful for degradation of cellulose or in generating anti-AviiiI antibodies for purifying recombinant AviiiI polypeptides from genetically engineered host cells.

Claim 3; Page 24; 44pp; English.

The present sequence encodes a thermostable avicelase polypeptide, designated AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes, and is a cellulase. AviiiI is useful in the conversion of biomass to biofuels and biofuel additives. It may be useful in the production of detergents, pulp and paper processing, food and feed processing and in textile processes. The thermostable AviiiI peptide is useful in the degradation of cellulose, and in generating specific anti-AviiiI antibodies that are useful in purifying recombinant AviiiI polypeptides from genetically engineered host cells, in detecting AviiiI polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the polypeptide. The AviiiI polynucleotide is useful as a source of probes or primers in various diagnostic assays

SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

## Alignment Scores:

Pred. No.: 9.08e-252 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 7 Gaps: 0

US-09-917-376-3 (1-740) x ABZ77632 (1-2869)

PA	XX	619	TGGTCCCGAGATGACGAACCTTTCCGGACGTAGCAGCTACATTGCAATCCCACTGACACG	678
PI	XX	181	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer	200
DR	XX	679	ACCGGCTATCAGAGCGATATTCAAGGCGTCTGCTGGGTCTTTCGACAAGTCTTCGTCA	738
XX	XX	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220
PT	XX	739	TCGCTCGGCGAAGCGAGTAAGACCAATTTTGTGGCGGTGGCGATCCCAATAATCCGCTC	798
PT	XX	221	PheTrpSerArgAspGlyValAlaThrTrpGlnAlaValProGlyValAlaProThrGlyPhe	240
XX	XX	799	TTCTGGACGAGACGGCGCGGACGTGGCAGCGGTGCGCGGTGCGCGGCGCTTC	858
PS	XX	241	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260
CC	XX	859	ATCCGCGACAGGGGCTTTGACCCGCTCAACACGCTCTATATTGCCACCAACAT	918
CC	XX	261	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly	280
CC	XX	919	ACGGGTGGTCCGTATGACGGAGCTCCGGCAGCTCTGGAATTTCTCGGTGACCTCCGG	978
CC	XX	281	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr	300
CC	XX	979	ACATGGACCGCAATCAGCCCGTACCTTCGACGACGCGCCCAACGACTACTTTGGTTAC	1038
CC	XX	301	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	320
XX	XX	1039	AGCGGCTCATCTATCGACCGCGCAGCACCGGACAGGATAATGGTGGCAACCCAGATATCG	1098
DB	XX	321	TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle	340
QY	XX	1099	TGGTGGCGGACACCAATAATCTTTCGAGACCGGCGGTGGCGAGCGCGGATC	1158
DB	XX	341	TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu	360
DB	XX	1159	TGGGATTGGACGAGTTATCCCAATCGAAGCTTGCATATGTCTTGACATTTTCGGCGGAG	1218
QY	XX	361	ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr	380
DB	XX	1219	CCTTGGCTGACTTGGGGTACAGCGGAATCTCCCGTACCAGTCCGAGCTCGGCTGG	1278
QY	XX	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly	400
DB	XX	1279	ATGGATGAAGCGATGGCAATCGATCCGTCAACTCTGATCGATGCTCTACGAAACAGGC	1338
QY	XX	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle	420
DB	XX	1339	GCGACGTTGTACGCAACAAATGATCTCAGAACTGGGACTCCGGCGCGCCAGATTCAATC	1398
QY	XX	421	AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer	440
DB	XX	1399	GCGCGATGGTCAAGAGATTGGAGAGCGGCTAAACGATCTCATAGCCCGCGCTCT	1458
QY	XX	441	GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
DB	XX	1459	GCGCGCGCTCATCAGCGCTCTCGAGACTCGGCGCTTCCACCCACCGACGCTTACT	1518
QY	XX	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
DB	XX	1519	GCCGTGCCATCGAGATCTTCACGTCACCGGTGTTCAGCACCGGACCGACCGCTCGACTAT	1578
QY	XX	481	AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro	500
DB	XX	1579	GCGAAATGAATTCGTTCGATCATCTTCGCTGGAGTTTCGATCTTCAGCCCAACCG	1638
QY	XX	501	AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGlu	520
DB	XX	1639	AACGACGACGCTCGGCTTCTCGACAGCGCGGCAAGAACTGGTCCCAAGCAGCGAA	1698
QY	XX	521	ProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheVal	540
DB	XX	1699	CCTGGCGGGTGACGACGGGCGGACCGCTCGCGCATCGGCGACGGCTCTCGTTCGTC	1758





Db 679 ACCGGCTATCAGAGCATATTTCAAGGCGTCTGTCGGTCTTTCGACAAAGTCTTCGTCA 738  
 Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
 Db 739 TCGCTCGGGCAGCGAGTAAAGCAATTTTGTGGCGTGGCGGATCCCAATAATCCGGTC 798  
 Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
 Db 799 TTTCTGAGCAGACGCGCGCGGACGTGGCAGCGGTGCGGGTGGCGGACCGGCTTC 858  
 Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuValIleAlaThrSerAsn 260  
 Db 859 ATCCCGCACAGGGCGCTTTGACCCGCTCAACACGCTCTATATTTGCCACAGCAAT 918  
 Qy 261 ThrGlyGlyProTrpAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
 Db 919 ACGGGTGGTTCGTATACGGGAGCTCCGGCGACGTCTGGAAATTCCTCGGTGACCTCCGG 978  
 Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
 Db 979 ACATGACCGGAATACAGCCGGTACCTTCGACGACACCGGCCAACGACTATTGGTTAC 1038  
 Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 Db 1039 AGCGGCTCACTATCGACCCGACGCCACCCGACACGATATGTGGCAACCCAGATATCG 1098  
 Qy 321 TrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
 Db 1099 TGGTGGCGGACACCAATATCTTTCGAGCACCGCGGGTGGCGGACGTGGCGGATC 1158  
 Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
 Db 1159 TGGGATGGAGGATTAATCCCAATCGAAGCTTCGATATGTGTCATTTGCGCGGAG 1218  
 Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTyr 380  
 Db 1219 CCTTGCTGACCTTCGGGTACAGCGAATCTCCCGTACCCAGTCCGAAGCTCGGCTGG 1278  
 Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 Db 1279 ATGGATGAGCGATGCAATCGATCGTTCACCTCTGATCGGATGCTCTACGGACAGGC 1338  
 Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 Db 1339 GCGACGTTGACGCAACAAATGATCTCACGAAGTGGGACTCCGCGCGCCAGATTTCATATC 1398  
 Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440  
 Db 1399 GCGCCGATGTCAGAGATTGGAGGAGACGCGGTAAACGATCTCATCAGCCCGCGCT 1458  
 Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
 Db 1459 GCGCGCCGCTCATCAGCGCTCTCGAGACCTCGGGGCTTACCCACGCGGAGTTACT 1518  
 Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480  
 Db 1519 GCGGTGCCATCGACGATCTTCAGCTCACCGGTTCACGACCGCCACCGCGTCCGACTAT 1578  
 Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
 Db 1579 GCGGAATGAATCCGTCGATCATCTTCGGCTGGGAAGTTTCGATCCATCGAGCAACCG 1638  
 Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520  
 Db 1639 AACGACAGGACGCTCGGCTCTCGACAGCGGGCGGCAAGACTGGTTCCNAGCAGCGAA 1698  
 Qy 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
 Db 1699 CTTGGCGGGTGACGACGCGCGGACCGTCCGCGCATCGCGCGCTCTCGTTTCGTC 1758  
 Qy 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560  
 Db 1759 TGGGCTCCCGGCGATCCCGTACGCTGTGTGTAGCGAGTTCGGATTTCGCAACTCTCTGG 1818

Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
 Db 1819 GCTGCTTCGCAAGGTGTTCCGCCAATCCCGAGATCCGCTCAGACCGGTGAATCAAAG 1878  
 Qy 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
 Db 1879 ACTTTCATGCCCTATCCAAATGGAACCTTCTATCGAAGCAGCGGCGGTGACATTC 1938  
 Qy 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620  
 Db 1939 CAACCGTTCGGCGCGGTCTTCCGAGCAGCGGTGCCGTGCTCATGTTCCACGCGGTG 1998  
 Qy 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640  
 Db 1999 CTTGAAAGAGAGCGGATCTGTGGCTGCTGCATCGAGCGGCTTTACCACTCAACCAAT 2058  
 Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
 Db 2059 GCGGGCAGCATGCTGCTGCAATCACCGCGGTATCTCCGCGTGAACGTGGGATTGTT 2118  
 Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal 680  
 Db 2119 AGTCTCGGCCGCGGTGCTCATACCCAGCCGCTTTGTCGCGCACGATCGAGGCGTT 2178  
 Qy 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700  
 Db 2179 ACGGGGCGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTGATCAATGATGACCAG 2238  
 Qy 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720  
 Db 2239 CACCAATACGGAATTTGGGCAAGCAATCACCGGTGACACCGGAATTTACGGCGGTG 2298  
 Qy 721 TyrIleGlyThrAsnGlyArgGlyValTyrGlyAspIleGlyValAlaProSerGly 740  
 Db 2299 TACATAGCAGAACGCGCGGTGGAATTTGATACGGGACATTTGTTGGTGGCCCGTCCGA 2358  
 RESULT 3  
 ABV76941  
 ID ABV76941 standard; DNA; 2823 BP.  
 XX  
 AC ABV76941;  
 XX  
 DT 03-MAR-2003 (first entry)  
 XX  
 DE Nucleotide sequence of a xyloglucanase enzyme.  
 XX  
 KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;  
 XX textile scouring; cellulose fiber processing; ratting; gene; ss.  
 XX  
 OS Jonesia sp.  
 XX  
 Key Location/Qualifiers  
 CDS 1..2823  
 FT /\*tag= a  
 FT /product= "xyloglucanase enzyme"  
 FT 1..69  
 FT /\*tag= b  
 FT 70..2820  
 FT /\*tag= c  
 XX  
 PN WO200277242-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-DK000210.  
 XX  
 PR 27-MAR-2001; 2001DK-00000504.  
 XX  
 PA (NOVO ) NOVOZYMES AS.  
 XX  
 PI Duffner F, Sjöholm C;  
 XX

DR WPI; 2003-092855/08.  
 DR P-PSDB; ABB99489.  
 XX New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,  
 PT and endogenous to a bacterium, useful in the textile industry for  
 PT improving properties of cellulosic fibers, yarn or fabric.  
 XX Claim 11; Page 68-72; 76pp; English.  
 XX The present sequence encodes a xyloglucanase enzyme, belonging to family  
 CC 74 of glycosyl hydrolases. The enzyme is isolated from *Jonesia* sp.  
 CC DSM14140. The enzyme is useful in processes for machine treatment of  
 CC fabrics. It is also useful in the textile industry for improving the  
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in  
 CC a textile scouring process step. The xyloglucanase enzyme is also useful  
 CC in the cellulose fiber processing industry for rattling of fibers such as  
 CC hemp, jute, flax and linen. It is useful for preventing binding of  
 CC certain soils to the xyloglucan left on the cellulosic material  
 XX  
 SQ Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3,88e-124 Length: 2823  
 Score: 2063.00 Matches: 386  
 Percent Similarity: 67.21% Conservative: 112  
 Best Local Similarity: 52.09% Mismatches: 225  
 Query Match: 51.11% Indels: 18  
 DB: 7 Gaps: 9

US-09-917-376-3 (1-740) x ABV76941 (1-2823)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25  
 DB 130 TATAGCTGTGCAAGCTTCAGATTGTGGCGGCGGTTTGTGCTGGATTGTCTTTAAC 189  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
 DB 190 CAGAAAGACCCAGGCTTGTGTATGTCGCCGACAGATATTGGTGGAGCTACCGGCTCAAT 249  
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 DB 250 GACTCCACCGGGGTGGATGCTCTTACGGACCATATTGGTTGGGATGACTGGAGCCAC 309  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 DB 310 TCCGGAATCTGTCTTGGCCACTGACCTGTGTGACACCAACCGCTGTACTTGGCAGCA 369  
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspGln 105  
 DB 370 GGCACGCTACTCCATGGACTGGATCCACAAACGGTGTCTATTGCTTCCGGCGGACAA 429  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
 DB 430 GGCAGACCTGGGAAAGACGATGCTGCGGCTCCGTGTGGAGGCAACATGCGCGGTGCG 489  
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
 DB 490 GGGATGGGAGACGCTCCCGGTGACCCCAACAATAATAAGCTCTGTACTTTGTGTGG 549  
 QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
 DB 550 GAATCGGGTATGGAATGGAAGTCACTGACTACGGGAAACCTGGGGCAAGTGACG 609  
 QY 166 AsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrGly---TyrGln 184  
 DB 610 AGCTTCCCAATGCTGGTAACCTACGTCGCG-----GATGCTTCTGGGTGCTACACC 660  
 QY 185 SerAspIleGlnClyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 204  
 DB 661 GGGCAAAACCAAGGCTTGTGGGTAACTTTTGACCCCACTACGCAAGAGCTGTGTGAG 720  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224

DB 721 ACAACACAAACGATCTACGTAGGTGTAGCAGACAGCAACAAAGCTTTATCGATCTACC 780  
 QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244  
 DB 781 GATGGCGGGCAACATGGCAACGGCTCCCTGGCAACCCACCGGATTCCTGGCGCAAAAG 840  
 QY 245 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
 DB 841 GGAGTGTCTTACCATAAAGGTGACCAACTCTACATCGCTACCTCAGATATGGTGGCCCG 900  
 QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpArg 284  
 DB 901 TATGACGGGTGCAAGGGGATGTGGCGCTAGATATTTCCAGTGGCCAGTGGACTCGA 960  
 QY 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304  
 DB 961 ATCAGCCCAATTCCTCTACC---TCATCGAATTCAGGTTCCGGTACTCCGGTTTAGCC 1017  
 QY 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
 DB 1018 ATTGACCGCAAAACCTGACACCATCATGGTTGTTCCCAAGTGTCTTGGTGGCTGAC 1077  
 QY 325 ThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThr 344  
 DB 1078 ATGTACGTATATCGTCCACCGACCGGGGGAACCGTGTGCTATCTGGAACTGAAC 1137  
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364  
 DB 1138 GGTTCACAGCGGTGTAACAGCAGTACACCATGATTACTCAGGGGCTCCCTGGCTTGAC 1197  
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384  
 DB 1198 TTTGGCAATCAGCAAAAGAGCCTGAAGCTAAACCCCAACCTCGGTGATGATGCCAGTCC 1257  
 QY 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404  
 DB 1258 TTCGAGATTACCCCGCATATACTCCGACGCTTCTTCTATGTCACCGCGCGGCTAT 1317  
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424  
 DB 1318 GCGGGAACGAACCTCAAAACTGGGTAAGGGCAAAAGTTGATATCCTGTCAAGGCC 1377  
 QY 425 LysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444  
 DB 1378 CAAGGATCGAAGAAACCGCAGCCCAAGACCTTCGCGCTCCCTCGGAAACATTTGGCTTA 1437  
 QY 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
 DB 1438 TACTCAGCGCTGGCAGACATCGGTGGTTCACCCCAAGATATCTCTCAGGTGCTTAAC 1497  
 QY 465 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483  
 DB 1498 AAGTACTACTACAAGAACCCACCATGACACCTGACCTCCATTCATTCGCGCAAGAC 1557  
 QY 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503  
 DB 1558 AAACCGCAACTGCTGCTGCTGAGGCAAA-----TCTATTCTGGGGAAACCACTCA 1611  
 QY 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
 DB 1612 TGGTGTGGGTTCACATGATGCTGGCGAAACGTTGAAACCTGGCGCCACGCAATCAGGG 1671  
 QY 524 ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543  
 DB 1672 GTTAAAGCCCGGCTCCATCCTGTCAGCTAATGCTTTCATCATTTGTGTGGCGCCCG 1731  
 QY 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563  
 DB 1732 -----GAAGTGGCGCTCCTCGAGCTTCCACCAACAGTGGCTCCTCGTGTCAACGGTG 1785  
 QY 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583  
 DB 1786 AGCGGGCTTCTCACAATGCCAGGTGCGCATCAGACCGAGTCAACGCTAAATCTCTTTTAC 1845

QY 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProVal 603  
 Db 1846 GGCTTCGTTACCGGAAGTCTATCACTCCACAAATGGTGGGCTTCTTCACCGCGTCA 1905  
 QY 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622  
 Db 1906 GCTTTCACGGGGTTCACACAGTGA-----AACGTTGTTCCGTGCGTCCCGGC 1959  
 QY 623 LysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637  
 Db 1960 CGTCAAGTTCATCTCTGGCTTGTGTGGATATCGGGAAGCACTATGGATGTGGCG 2019  
 QY 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657  
 Db 2020 TCCACTGATCGCGGAAGAACTGACCAAGTATCTCGAGTTTCAGGAAGCGATGCCGTT 2079  
 QY 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677  
 Db 2080 GGTTCGTAAGCAACAGTAGTAGCGGATACCACTGATTTTCATCCGCGCAAGATT 2139  
 QY 678 GlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeuIleAsn 697  
 Db 2140 GATGGATTTCAGGAATTTTCGGTCTGACGATGAGCGAAGACCTGGAACGCATCAAC 2199  
 QY 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717  
 Db 2200 GACAACCAACACAGTGTATTGACTGCGGCATCAATCCCGTGACCCAGATGTCGAA 2259  
 QY 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737  
 Db 2260 GGGCGGCTTACATTGGTACCAACGGTGGCGGATATCGTTGGTGTGATTCCTCCACTCCC 2319  
 QY 738 Pro 738  
 Db 2320 CCG 2322  
 RESULT 4  
 ADD42060  
 ID ADD42060 standard; DNA; 2849 BP.  
 XX AC ADD42060;  
 XX AC  
 XX 15-JAN-2004 (first entry)  
 XX DT  
 XX DE  
 XX Trichoderma reesei Family 74 xyloglucanase gene, SEQ ID NO:1 #2.  
 KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
 KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
 KW ethanol production; detergent composition; fabric treatment;  
 KW textile treatment; gene; ds.  
 XX OS Hypocrea jecorina.  
 XX FH  
 XX Key Location/Qualifiers  
 CDS 59..2743  
 FT /\*tag= b  
 FT /product= "Family 74 xyloglucanase"  
 FT /note= "Contains 3 introns"  
 FT 59..381  
 FT /\*tag= a  
 FT /number= 1  
 FT 382..440  
 FT /\*tag= c  
 FT /number= 1  
 FT 441..1227  
 FT /\*tag= d  
 FT /number= 2  
 FT 1228..1277  
 FT /\*tag= e  
 FT /number= 2  
 FT 1278..2664  
 FT /\*tag= f

FT intron  
 FT 2665..2723  
 FT /\*tag= g  
 FT /number= 3  
 FT 2724..2743  
 FT /\*tag= h  
 FT /number= 4  
 XX  
 FN WO2003089598-A2.  
 XX  
 XX 30-OCT-2003.  
 XX  
 PF 17-APR-2003; 2003WO-US011831.  
 XX  
 PR 19-APR-2002; 2002US-0373987P.  
 XX  
 XX (NOVO) NOVOZYMES BIOTECH INC.  
 XX  
 XX Michael R, Zaretsky E, Haas J;  
 PI  
 XX WPI; 2003-845528/78.  
 DR  
 XX P-PSDB; ADD42055.  
 XX  
 PT New polypeptides having Family 74 xyloglucanase activity, and encoding  
 PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
 PT -containing biomass to ethanol or as a detergent.  
 XX  
 PS Claim 1; Fig 1; 96pp; English.  
 XX  
 CC The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
 CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
 CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
 CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
 CC The invention also relates polypeptide sequences at least 70% identical  
 CC to the enzyme, expression vectors and host cells comprising a nucleic  
 CC acid of the invention, the recombinant production of the enzyme, and  
 CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
 CC the invention can be used in the degradation of cellulose- and  
 CC hemicellulose-containing biomass to produce ethanol. It can also be used  
 CC in a detergent composition for treating fabric during a machine washing  
 CC cycle. The present sequence represents the gene encoding Trichoderma  
 CC reesei Family 74 xyloglucanase. Note: A comparison of this sequence with  
 CC that shown in the sequence listing (also referred to as SEQ ID NO:1)  
 CC shows that the present sequence is half the length of the sequence  
 CC listing SEQ ID NO:1 (ADD42054), as the sequence listing SEQ ID NO:1  
 CC contains alternate 60 bp blocks of sense strand and complementary strand  
 CC sequence.  
 XX  
 SQ Sequence 2849 BP; 594 A; 882 C; 822 G; 551 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,23e-88 Length: 2849  
 Score: 1506.00 Matches: 329  
 Percent Similarity: 56.60% Conservative: 121  
 Best Local Similarity: 41.38% Mismatches: 260  
 Query Match: 37.31% Indels: 87  
 DB: 9 Gaps: 23  
 US-09-917-376-3 (1-740) x ADD42060 (1-2849)

QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 Db 119 TTTTCATGGAAGAACGTCAAGTCGGCGCGCGGCTTCTGCCCGCATCATCTTC 178  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
 Db 179 CATCCCAAGACAAAGGCGTAGCATATGCACACAGATATTGGCGGCTGTACCGCTC 238  
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
 Db 239 AAC---GCCGACGACTCATGACCGCCCTGACGGATGGATTCGTGATTAATCCCGCTCG 295  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80



## RESULT 5

ADD24893 standard; DNA; 3668 BP.

XX AC ADD24893;

XX DT 15-JAN-2004 (first entry)

XX DNA encoding Xanthomonas campestris cellulase #1.

XX Directed genetic engineering; galactomannanase; reduced activity;  
 XX enhanced activity; xanthan gum production; suspension stability;  
 XX emulsion stability; temperature resistance; pseudoplasticity; amylose;  
 XX cellulase; extracellular protease; intracellular protease;  
 XX glucose dehydrogenase; gene; ds.

XX OS Xanthomonas campestris.

XX PN US2003036176-A1.

XX PD 20-FEB-2003.

XX PF 10-AUG-2001; 2001US-00927827.

XX PR 28-MAR-2001; 2001US-0279493P.

XX PA (BOWE/) BOWER S G.

XX PA (RAMS/) RAMSEIER T M.

XX PI Bower SG, Ramseier TW;

XX DR WPI; 2003-625389/59.

XX DR P-PSDB; ADD24919.

XX PT New transformed cell or organism having reduced or enhanced activity of  
 XX at least one protein, useful for producing xanthan gum, which are useful  
 XX for providing formulations and properties.

XX Claim 15; SEQ ID NO 21; 135pp; English.

XX The present invention relates to polypeptide and polynucleotide sequences  
 XX from Xanthomonas campestris which may be used for activity reduction or  
 XX enhancement using directed genetic engineering. A transformed cell or  
 XX organism having reduced or enhanced activity of at least one such protein  
 XX e.g. galactomannanase can be generated by disrupting the gene encoding  
 XX the protein. The activity of the protein is reduced by the presence of an  
 XX antisense nucleic acid sequence. The nucleic acid sequence of the gene  
 XX encoding the protein is a recombinant sequence having at least one  
 XX mutation as compared to the wild-type gene encoding the protein. The  
 XX transgenic cell or microorganism are useful for producing xanthan gum,  
 XX which are useful for providing formulations and properties, such as long-  
 XX term suspension and emulsion stability in alkaline, acid, and salt  
 XX solutions, temperature resistance, and pseudoplasticity. The present  
 XX sequence encodes an enzyme relating to the present invention.

XX SQ Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3.13e-64 Length: 3668  
 Score: 1139.00 Matches: 267  
 Percent Similarity: 50.85% Conservative: 120  
 Best Local Similarity: 35.09% Mismatches: 295  
 Query Match: 28.22% Indels: 79  
 DB: 9 Gaps: 21

US-09-917-376-3 (1-740) x ADD24893 (1-3668)

QY 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20

Db 1217 GCCACGTCGGGGCCCTACCAAGTCGCGAGTGCCTATTCGGCGTGGCGCTTGTCCACC 1276

QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40

Db 1277 GGTGTCTGTTTCATCCCGCGCAACGTCGTCTGGCCATATGCGCGCACCGCATGTGGGTGGC 1336  
 QY 41 MetTyrArgTrpAspAlaAlaAsnGlyYArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db 1337 GCGTACCGCTGGGATGCGCAGCGCAGTGGACCGCGCTGACCGACTGGTGGCGCT 1396  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 Db 1397 GACGACTGGAACTGATGGGCATCGACGATTCGCGTCGACCCCGCCGATCCGATCG 1456  
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 Db 1457 CTGTATCTGGCGCGCGCACCTATATGATGAA---CGCGCGCGCACTGCGCGAGTGTG 1513  
 QY 101 ArgSerSerAspGlnGlyAlaValThrTrpGlnIleThrProLeuProPheLysLeuGly 120  
 Db 1514 CGCTCGTTCAACCGCGCGCGCACGTTGAGCGTGCAGCTTGCCTTTAAGTGGGTGGT 1573  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 Db 1574 AACGAGCTGGCGCGCGCAATGGCGAGCGGTGGCGTGGACCCCGCACGATGGCGCGGTG 1633  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
 Db 1634 CTGCTGTGGGCTCGCGCGATGCC---GGCCTGTGGCGTAGCGAGTATCGCGCGCGCAC 1690  
 QY 161 TrpSerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
 Db 1691 TGGCGGAAGTGGCGTCTGTTTCGGACCGCGCGTGGCGGTGCCACCGCGCGCAATCAT 1750  
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
 Db 1751 GTTGGCGCGAGCAGCGGTG---GGGATCGCTTTGTGTCGTTCGACGACGACGAGTGGC 1807  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
 Db 1808 AACAAACGGCTCGCAACACCGCGCATCTACGTGGCGGTGCCACCGAACAGACGACCTG 1867  
 QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValAlaProThrGlyPhe 240  
 Db 1868 TATGTGTCGGAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1927  
 QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
 Db 1928 CGCCCGAGCCACATGGCG 1984  
 QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
 Db 1985 CAGCCCG 2044  
 QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
 Db 2045 CGCTGGCGTGAGATCAGCCCGCATTCGCGAG---CCAGCCAGTGGCGGTGATGCTGCTG 2101  
 QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 Db 2102 GGTGCGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTGGCGCGCGCGCGCGCGCT 2161  
 QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
 Db 2162 CGCAGCG 2221  
 QY 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
 Db 2222 -----TTGCGCGATGCGGTGTTTCGATCATCAGACGCGCGCG 2254  
 QY 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380  
 Db 2255 CCGTGG---ACCGCACATGCCACCGCGCAC-----TGG 2284  
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400

Db 2285 ATGGGG---CGCGTGGCGATCGATCGTTCCAGCGCAACCATCGGCTGTTCTCGTACCGGC 2341  
Qy 401 AlaThrLeuTyxAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 2342 TACGGCATCTGGGCTCGCGCAATCTG-----CAGGATTTCCGCC 2380  
Qy 421 AlaPro-----MetValIysGlyLeuGluGluThrAlaVal 432  
Db 2381 GCACCGCAGCGCCGCTGCAGTGGTTCACGACCGTGGCTGGAAGAAACCGTCCCG 2440  
Qy 433 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
Db 2441 CTGAGCTCTCAGCCCGATGGCTGGCGCATCTGCTACGCGCTCGCGCATATCGAC 2500  
Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472  
Db 2501 GGCTTCGGCATCAGCACCTCGACCGGTG---CAGTTGACGATACCGCGCCACGGCTG 2557  
Qy 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492  
Db 2558 ACCAATGGCGAAGCATCATGATCGCGCGCGCAGCGCGCGTGGTGGTGGCGCGGT 2617  
Qy 493 SerPheAspProSerSerClnProAsnAspArgHisValAlaPheSerThrAspGlyGI 512  
Db 2618 ACCGTGCGCGCGCGCGCGCAACAAATCCGCGCGC---TGTATTACGCGTGTATGGCG 2674  
Qy 512 YLysAsnTTPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531  
Db 2675 CAGCATGAGATCGCTTTGCGAGTGAGCGCCGCGCGCGCGCGCGCGCGCATTCG 2734  
Qy 531 aAlaSerAlaAspGlySerArgPheValTnPAlaProGlyAspProGlyGlnProValva 551  
Db 2735 CATTTGGTCCGATCGCGCGCAGGTGGTGGCGCACCGCGCGCGCGCGCAAT-----TG 2788  
Qy 551 lTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGI 571  
Db 2789 GCGCAGCTCCGACTTCGCGCGCAGTGGCGACGCGTGTGAACGCGCTGCGCAACCCCGCGT 2848  
Qy 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589  
Db 2849 GGTGGTGCAGATCGGTGGTGCAGCAGCGCGCTGGTATGCGGTGTATCGCGCGCGGCA 2908  
Qy 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu----- 607  
Db 2909 GCTGTACGAGACCGATCCCGCGCGCAGTTCGT-----GCGACCGGTGTACAGGT 2962  
Qy 608 -----ProSerSerGlyAlaVa 613  
Db 2963 GGGCAGCGCGCGCGATGAGCGCACCGCGCGCAGCTGCTCCGACCGCGTGGCGCGC 3022  
Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633  
Db 3023 AGCGTGGTGTATCTGGCGACCGCGCGCAAG----- 3053  
Qy 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653  
Db 3054 -GGTGTATGCGCTGGCAGACCGT-----GCCTGACGCTGCTCTACAGCGCGGACGA 3106  
Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673  
Db 3107 AGCAGCTCGCTGGCATCGCAAGCGCGTGGCGCGCGTGGCGCGCGTGTATCT 3166  
Qy 673 lValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr 693  
Db 3167 GCGCGCGCGCTGACGAGCGTGTATGCGGTTCCTCGCTCCGACGATGCGCGCGCGAGTG 3226  
Qy 693 pValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713  
Db 3227 GCAGCGCATCAAGATGACCGCGCACCGCTTCGGCGCG---CCGTACAGCGTACCGGTGA 3283  
Qy 713 pHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733  
Db 3284 TCCGCGCATTCGCGCGGTGTGTACTTCGCGCACCGCGCGCGCGCATTTTCTACGCGCA 3343

Qy 733 p 733  
Db 3344 T 3344

## RESULT 6

ID ADD42054 standard; DNA; 5698 BP.

AC ADD42054;

DT 15-JAN-2004 (first entry)

XX Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.

XX Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
KW ethanol production; detergent composition; fabric treatment;  
KW textile treatment; ds.

OS Hypocrea jecorina.

XX WO2003089598-A2.

XX 30-OCT-2003.

XX 17-APR-2003; 2003WO-US011831.

XX 19-APR-2002; 2002US-0373987P.

XX (NOVO ) NOVOTYMS BIOTECH INC.

XX Michael R, Zaretsky E, Haas J;

XX WPI; 2003-845528/78.

XX P-PSDB; ADD42055.

PT New polypeptides having Family 74 xyloglucanase activity, and encoding  
nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
-containing biomass to ethanol or as a detergent.

XX Claim 1; SEQ ID NO 1; 96pp; English.

XX The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
The invention also relates polypeptide sequences at least 70% identical  
to the enzyme, expression vectors and host cells comprising a nucleic  
acid of the invention, the recombinant production of the enzyme, and  
mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
the invention can be used in the degradation of cellulose- and  
hemicellulose-containing biomass to produce ethanol. It can also be used  
in a detergent composition for treating fabric during a machine washing  
cycle. The present sequence represents a Trichoderma reesei Family 74  
xyloglucanase-encoding DNA. Note: The present sequence (referred to as  
SEQ ID NO:1) is given as shown in the sequence listing. A comparison of  
this sequence with that shown in figure 1 (also referred to as SEQ ID  
NO:1) shows that the present sequence is twice as long as the figure 1  
sequence (ADD42060) and contains alternate 60 bp blocks of sense strand  
and complementary strand sequence.

XX Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3.69e-47 Length: 5698  
Score: 878.50 Matches: 334  
Percent Similarity: 29.92% Conservative: 119  
Best Local Similarity: 22.06% Mismatches: 270  
Query Match: 21.77% Indels: 793  
DB: 9 Gaps: 52

US-09-917-376-3 (1-740) x ADD42054 (1-5698)





```
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTCTTAACGATCCGATGCGAATGG 2459
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370
Db 2460 CTGACTCTGGATAATGATGTGTAGAGTCACCTCAGTGAGAAATGTGTAGGCTAGCTTACC 2519
QY 371 -----ProProValPro----- 374
Db 2520 GACTGACTCTGCTTCAAGACTCCAAAGCACCGTGGATCAAGAAACAACCTTTATCGATGTG 2579
QY 374 ----- 374
Db 2580 ATGACTGACACGAAGTTCTCAGGGTTTCGTGGCACCTAGTCTTGTGTAATAGCTTACAC 2639
QY 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385
Db 2640 TCAGCGAGTCAACCGTCCGATGCTCTCAAGCGCTCGGCTGGATGATTGAGTCTCTC 2699
QY 386 Ala----- 386
Db 2700 GGCTCGCTCAGTGGCAGGTACACAGTAGTTTCGGGAGCGGACCTACTAACTCAGAGAG 2759
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
Db 2760 CAGATTGACCAACACGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGC 2819
QY 406 Thr----- 406
Db 2820 GTCTAACTGGGTTGGCTGTGTTGTGTGACCGAGATCCGTGGCTTACTGTTAGAAACCG 2879
QY 407 ---AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425
Db 2880 CGCCAGACTCACCAACTGGGACACGGCCACAATGTGTCATCAATCAATCACTGGCAGAC 2939
QY 426 Gly----- 426
Db 2940 GCGGTCTAGAGTGGTTGACCTGTGCGCGGTGTTACACAGTTAGGTTAGTGCAGCTGTG 2999
QY 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIle 445
Db 3000 CGCATCGAGGAATTTCTCCGTCCAGGACCTGGCTCTGCAACCGCGGAGCGAGCTATTG 3059
QY 446 Ser----- 446
Db 3060 GCGTAGCTCCTTAAGAGGAGGTCTCGACCGGAGACGTGGCGCGCTTCGCTCGATAAC 3119
QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462
Db 3120 CCGCGAGTCGAGACGACACACGCTTCACTTTGCCAGCAGAAACGACCTCGGGACATCG 3179
QY 462 ----- 462
Db 3180 CGCGGTGAGCTCTGCTGTTGCCAAGTGGAAACGGTGTCTTTGCTGGAGCCCTGTAGC 3239
QY 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla--- 481
Db 3240 GCGCAGACGCTCTGGGCAACGCCCCACATGGGCCACCTCGACGACGCGTCTGACGCGGG 3299
QY 482 -----Glu 482
Db 3300 AGGCTGTCCAGACCCGTTGCGGGTGTACCCGGTGGAGTGTCTGCGAGCTGATCGGCCCC 3359
QY 483 LeuAsnProSerIleValArgAlaGlySer----- 493
Db 3360 TACTCGGTCAAGACGTCGTCCGCGTGGCAACACACCGCGGACGCAACAGGTGGCCATC 3419
QY 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPhe----- 507
Db 3420 TTGAGCCAGTTCTTCGACGAGGCGGACCGCTTGTGGCGCGCGTGTGTCACCGGTAG 3479
QY 508 SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrGly 527
Db 3480 ACGTCCGACGGCGGCGGACGCTGGAGCATCGACTACGGGCCGACACGTCATGAACGCG 3539
QY 528 Gly----- 528
Db 3540 GGCAGGCTGCGCGCGCTGCACCTCTAGTGTGATGCGCGGCTGTGAGGTACTGTGCG 3599
QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547
Db 3600 CGCACGGTGGCTATTTCGGCGGACGCGGACACGATCCTCTGTGTGACCGCCTCTGTCGCG 3659
QY 548 -----GlnPro 549
Db 3660 GCGTCCACCGGATAACCGGCTGCCGCTGTGCTAGGAGACCAGTGGCGGAGCAGGCG 3719
QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569
Db 3720 CTCGAGCGCTCG---CAGTTCCAGGCGAGCTTTTGCTCTCGAGCTTGCCCGCGGCG 3776
QY 570 AlaGln----- 571
Db 3777 GCGGACGTCGCGAGCGTCAAGGTCCTCGAAACGAGGACGAGCTCGGACGCGGCGCG 3836
QY 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589
Db 3837 CGGCTCATCGCTCGGACAAAGAACCAACAGCGTCTTCTACCGCGGCTCGGATCGACC 3896
QY 590 Phe----- 590
Db 3897 TTTTAGTAGCGGACCTGTTCTTCTGTTGTCGAGAAGATGCGGCCGAGGCTAGCTGG 3956
QY 591 ---TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609
Db 3957 AAAAACTGACGAAGGACACCGGCGAGCAGCTTC-----ACGCGCGGG---CCCAAG 4004
QY 610 SerGlyAlaValGlyValMet----- 616
Db 4005 CTGGGACGCGAGGATGCACTGTTCTGTGGCGCTGTCGAAAGTGGCGGCCGCGGTTTC 4064
QY 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627
Db 4065 GACCCGTGCGGTCTCCGTCCGATCCGGGATATCGCTCTACCCGACACCGCGGCGAGTTG 4124
QY 628 TrpLeuAlaAlaSerSer----- 633
Db 4125 TATGCTCGACCGCGGTAGGCCCTATAGCAGAGTGGCTGTGTGGCGCCGTCGAAC 4184
QY 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644
Db 4185 ATACAGAGCTGGCTGTCTCGGCATATTCGCTCCACAGACTCGGGCAGCACCTTTGGCCAA 4244
QY 644 ----- 644
Db 4245 GTCTCCACCGCCTGAACCGGTATAAGCGAGGTGTCTGAGCCCGCTGTGGAACCGGTT 4304
QY 645 ---TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 4305 CAGAGTGGCGGACTCCACACCTACAGATCGCCCTGGGTGTGGCTCAGCTCGAAC 4364
QY 663 -----AlaProGlySerSerTyr 668
Db 4365 TGGAACTGTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4424
QY 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyVala-----TyrArgSer 686
Db 4425 ACCTTGGACATACGATCGGCACC-----GGCCCGTCAAGGGCTCGCCTCTACGCCAGT 4478
QY 687 AspAspCysGlyThrThrTrpVal----- 694
Db 4479 GGAGACAGCGCGCTCTCTGGAAGCCGTGGCGGCGAGTCCCGGAGCGGAGATGCGGTCA 4538
QY 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly----- 704
Db 4539 CCTCTGCGCGGAGGAGACCTCGGACATCGAGGCTCCAGGGCTCCAGGGCTTCGCTCCATCGAC 4598
```

Qy 704 ----- 704  
 Db 4599 AGCACCAGTCCGCGCAGCGGCTGTAGTCCCGAGGGTCCCGAAGCCGAGGTAGCTG 4658  
 Qy 705 AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThr 724  
 Db 4659 TCGTGGTTCAGCGCGCGCTCGCGAGCACCAGCGCGG-----CAAGTCTACGTGGGCACC 4712  
 Qy 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
 Db 4713 AACGGCGCGCGCTTTTACGCTCAGCGCTCGTGGCGGCC 4754  
 RESULT 7  
 AAF14988  
 ID AAF14988 standard; cDNA; 1103 BP.  
 XX  
 AC AAF14988;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 13-MAR-2001 (first entry)  
 XX  
 Trichoderma reesei EST SEQ ID NO:7511.  
 XX  
 Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Hypocrea jecorina.  
 XX  
 WO200056762-A2.  
 XX  
 28-SEP-2000.  
 XX  
 22-MAR-2000; 2000WO-US007781.  
 XX  
 22-MAR-1999; 99US-00273623.  
 XX  
 (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 PA  
 Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 PI WPI; 2000-594572/56.  
 XX  
 Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 Claim 89; Page 3034; 3161pp; English.  
 PS  
 XX  
 The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered.  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organization of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents

CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from  
 CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
 CC Trichoderma reesei, which are all specifically claimed in the present  
 CC invention. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 0 U; 21 Other;

# Alignment Scores:

Pred. No.: 1.54e-16 Length: 1103  
 Score: 392.50 Matches: 114  
 Percent Similarity: 45.05% Conservative: 36  
 Best Local Similarity: 9.72% Mismatches: 125  
 Query Match: 3 Indels: 58  
 DB: 3 Gaps: 10

US-09-917-376-3 (1-740) x AAF14988 (1-1103)

Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321  
 Db 3 GGCCTTGGCTCGATTTGCAAAAGCCAGGACCCCTTGTGTGCTTCTTTGAACCTCTGG 62  
 Qy 322 TrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341  
 Db 63 TGGCAGATGCTCAGCTGTTTCGGTCGACCACTCTGGGACCAACATGGAGCCGATCTGG 122  
 Qy 342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361  
 Db 123 CGGTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAGCACCG 182  
 Qy 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 376  
 Db 183 TGGATCAAGAACAACTTTTATCGATGTGACGAGCGAGTCACCGTCCGATGCTCTNATCAAG 242  
 Qy 377 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396  
 Db 243 CGCTCGGCTCGATGATTGAGTCTNCTGAGATTGACCCACCGCAGCAANNACTCGCTT 302  
 Qy 397 -TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer----- 414  
 Db 303 TTACGGCACCGAATGACAAATNTTGGCGGCCACGATTTCCACCACTGGGACACGGCCCC 362  
 Qy 415 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluGluThrAlaV 432  
 Db 363 ACAATGTGGTCAATCCCAATTACTTGGCAGACGGGATTCGAAGGAATT-----TTTCCGT 416  
 Qy 432 alaAsnAspLeuLeuSerProProSerGly----- 441  
 Db 417 TCAAGGACCTGGCGCTTTTACCCCGGGGGAAGCGAGCTTTTGGCCGCAAGTCGGGAGACG 476  
 Qy 442 -----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV 459  
 Db 477 ANCAACGGGTTTACCTTTTGGCCAGCAAGACCTTTGGG----- 517  
 Qy 459 alThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr-SerVal 478  
 Db 518 -----ACATTCGCCGACAGCGGTTTGGGCACTCCACATGGGCCACCTCGACGAAGCGTC 572  
 Qy 479 AspTyrAla--GluLeuAsnProSerIleIleValArgAlaGlySer----- 493  
 Db 573 GACTACCGCGGAACCTCGGTCAAGAGCGCGTTCGTCGCGTCGCGCAACACCGCCGCGCAGC 632  
 Qy 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 508  
 Db 633 CAACAAGGTGCCATTTTTCGCAACGGCGGGCGCCGACGTNGAA-----GCAATTC 686  
 Qy 508 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG 528  
 Db 687 GAACATAACGCT-----GGTCCGAACACCGTTTCCATTCAACACGGCG 728  
 Qy 528 lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 548  
 Db 729 GCGCGGTGGCTATTTCGCGCCGACGACGACGATCTCTGTCGACCGCGCTCGTCCGGC- 787

QY 548 lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 568  
 Db 788 -----GTGAGCGCTCGAGTTCAGGGCAGCTTTCCTCGTCTCGAGCTCGCCG 839  
 QY 568 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG 588  
 Db 840 CGGGCGCGTCTCGTCTCGGCAAGAACACACGAGCGTCTTCTAGCGCGTCTCGGAT 899  
 QY 588 lyThrPheTyrArgSerThrAspGlyGly 597  
 Db 900 CGACCTTTTACGTACAGCAAGGACACCGGC 928

## RESULT 8

ACA42281  
 ID ACA42281 standard; DNA; 7407 BP.

AC ACA42281;  
 DT 19-JUN-2003 (first entry)  
 DE Prokaryotic essential gene #23938.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.

XX Pseudomonas aeruginosa.

OS WO200277183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

XX P-PSDB; ABUS8411.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 30151; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 7407 BP; 1387 A; 2763 C; 2378 G; 879 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.36e-05 Length: 7407  
 Score: 238.00 Matches: 205  
 Percent Similarity: 32.59% Conservative: 88  
 Best Local Similarity: 22.80% Mismatches: 314  
 Query Match: 5.90% Indels: 292  
 DB: 7 Gaps: 50

US-09-917-376-3 (1-740) x ACA42281 (1-7407)

QY 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30  
 Db 3250 GTCAACCTGAGCAATGCGAGCAGCCTCAGCGC-----ACGCGCGAGCGGCG 3297  
 QY 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
 Db 3298 AGCACGCTGATTCTCACCGAC-----GGCAACGCGCAAT 3330  
 QY 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr-----65  
 Db 3331 CCGATCCCGGAGTACCGCGCAGCGAGCGGCACTGAGCACTACACCCGTCACGCGG 3390  
 QY 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
 Db 3391 ATCGCAACGCTACTGTGGTCAACGTCGTCGCCAGGAGCGCTCCGGTAACAGCAGCGG 3450  
 QY 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
 Db 3451 CCGCGCAGCGTGACCGTCTGATTTCAGCGCGCGCGCGCGCGCGGTGATCAACCCGAGCAAC 3510  
 QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
 Db 3511 GCGTGTGTATCAGCGCGCAGCGCGCGCGCGCGCGCGGTGATCAACCCGAGCAAC 3561  
 QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135  
 Db 3562 ---GATCCGCGCGGCAACCCGATAGGCGAGTACCGCGCGAC-----3600  
 QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
 Db 3601 -----GGCAGCGGCAACTGGAGCTTCACG 3624  
 QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175  
 Db 3625 CCG-----GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3642  
 QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
 Db 3643 AACGCGACGCTGATCTGCGCCACCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3702  
 QY 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201  
 Db 3703 GCCACACCGTGGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3762  
 QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
 Db 3763 ACCATCAGCGCGCACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3822  
 QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237



XX 22-MAY-2003.  
 XX PD  
 XX PF 18-SEP-2002; 2002WO-US029565.  
 XX PR 18-SEP-2001; 2001US-0323241P.  
 XX PA (DART-) DARTMOUTH COLLEGE.  
 XX PI O'toole GA, Mah T;  
 XX DR WPI; 2003-468567/44.  
 XX DR P-PSDB; ABP59933.  
 XX PT Identifying modulators of microbial resistance of organisms in biofilms,  
 PT e.g. inhibitor of biofilm formation, by employing expression controls, or  
 PT efflux pumps containing polypeptides, of genes associated with biofilm  
 PT resistance.  
 XX PS Disclosure; Fig 7; 102pp; English.  
 XX CC The present invention relates to a method of identifying a compound  
 CC capable of altering the sensitivity of a microorganism to an  
 CC antimicrobial agent by employing efflux pumps comprising polypeptides  
 CC encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143,  
 CC PA2390 or PA1163. The method is useful for identifying modulators of  
 CC microbial resistance of an organism in a biofilm. The methods are also  
 CC useful for identifying genes that encode proteins that play a role in  
 CC biofilm resistance. The method is particularly useful for screening  
 CC compounds or discovering compositions that will inhibit biofilm formation  
 CC and overcome their resistance mechanisms. These methods are particularly  
 CC useful in medical, industrial or natural settings, where formation of  
 CC biofilms can have serious negative consequences and result in high costs  
 CC both in human health and economic terms. The present sequence is a  
 CC coding/control sequence shown in the exemplification of the invention  
 XX SQ Sequence 7407 BP; 1387 A; 2763 C; 2378 G; 879 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.36e-05 Length: 7407  
 Score: 238.00 Matches: 205  
 Percent Similarity: 32.59% Conservative: 88  
 Best Local Similarity: 22.80% Mismatches: 314  
 Query Match: 5.90% Indels: 292  
 DB: 8 Gaps: 50

US-09-917-376-3 (1-740) x ACC59398 (1-7407)

QY 11 ValAlaIleGlyGlyGlyPheValAspGlyValPheAsnGluGlyAlaProGly 30  
 DB 3250 GTCAACCTGAGCAATGGCAGCAGCTCAGCGGC-----ACCGCGAGCGCGC 3297  
 QY 31 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
 DB 3298 AGCAGCGTGATCTCACCGAC-----GGCAACGGCAAT 3330  
 QY 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr-----65  
 DB 3331 CCATCGCGAGGTTCACCGCGCAGCGGCACTGGACCTACACCCCGTCCACCGCG 3390  
 QY 66 -----AsnGly-----ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
 DB 3391 ATGCCCAACGGTACTGTGTGTCAACGTGTGGCCAGCAGCGCTCCGGTAAACAGCAGCCG 3450  
 QY 83 AlaAlaIleValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
 DB 3451 CCGCGCAGCGGTGACCGTCGATTCCAGCGCGCGCGCGGTGTGATCAACCGCAGCAAC 3510  
 QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
 DB 3511 GCGGTGTTCATCAGCGGCACCGCGCGCGGTGTCCACCGTGCACCTCACC-----3561  
 QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135

DB 3562 ---GATCCCGCGCAACCCGATAGGCGAGCTACCGCGCAG-----3600  
 QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
 DB 3601 -----GGCAGCGGCAACTGGAGCTTCAGC 3624  
 QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175  
 DB 3625 CCG-----GGCACCCCGCTGGCC 3642  
 QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
 DB 3643 AACGCGCAGGTGATCGTCGCCACGCCACCGCAGCGCAATACCGCGCCCGCAGCGC 3702  
 QY 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201  
 DB 3703 GCACACACGGTGGACGGGTGGCGCGCGCGGTGTGATCGATCCGAGCAACGGCAGC 3762  
 QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
 DB 3763 ACCATCAGCGGCACCGCGGAGCGCGGCAAGGTGATCTCACCGCAGCGCAACGGCAAC 3822  
 QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
 DB 3823 CCGATCGCGGAAACACCGCGCGCAGCGCAACTGGAGCTTCACGCCCGCAGCGCGC 3882  
 QY 238 -----ThrGly---239  
 DB 3883 CTGGCCAAACGGCAGGTGGTCAACCGCGTGGCCCGCAGGACCTCGGGGCAATACCGCGCGC 3942  
 QY 240 -----PheIleProHisLysGlyValPheAspProValAsn 251  
 DB 3943 CAGGGCAGCACTACCGTGGACCGGTGGCGCGGCAACAGCGCTGTGTCAATCCGAGCAAC 4002  
 QY 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268  
 DB 4003 GGCAACCTGTCAACGGTACCGCGGAGCGCGGAGCACCGTGACCTTCAGCGCAGCGCAAC 4062  
 QY 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285  
 DB 4063 GGCAACCCGATCGGCCACCGCAGCGCGCAGCGCACTGGAGCTTCACCGCGCGC 4122  
 QY 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
 DB 4123 TCGCAACTGCCCAACGGCAGCGGTGTCAACGTGACCGGAGCGCGCGCGGCAATACC 4182  
 QY 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal-----315  
 DB 4183 AGCTTCCCGTACCGCAGCGGTGATTCCTCGTGGCGTGCATCCCGCAGGTGGATCCG 4242  
 QY 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330  
 DB 4243 AGCAACGGTTCGGTGTATCAGCGGCACCGCGCAGCGCGCAACCATCATCATC-----4296  
 QY 331 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAsp 342  
 DB 4297 ACCGATGGCAACGGCAACCGGATGGCGCAGGTACCGCGCGCAGCGCGGCACTGGTCC 4356  
 QY 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354  
 DB 4357 TTCCTCCAGGCATCCCGTGGCGGATGGCAGCGGTGTCAACGTGGTGGCGCGCAGCCCA 4416  
 QY 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnProPro 372  
 DB 4417 AGCAATGTGCACAGTGGCGCGCGGTGATCACTGTGTGTGGTGGCGCGCGCGCGCGC 4476  
 QY 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
 DB 4477 GTG-----ATCGATCCCGAGCAACGGC 4497  
 QY 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409





CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 7305 BP; 1395 A; 2299 C; 2566 G; 1045 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0.00414 Length: 7305  
 Score: 199.50 Matches: 185  
 Percent Similarity: 32.00% Conservatives: 72  
 Best Local Similarity: 23.04% Mismatches: 303  
 Query Match: 4.94% Indels: 243  
 DB: 7 Gaps: 42

US-09-917-376-3 (1-740) x ACA23399 (1-7305)

QY 47 AlaAsnGlyValTrp-----IlePro 53  
 DB 517 GCGACGGCGCTGTCGTACAGCCCGGTACGTCCTCCGAGGCGCCCATCAGATACC 576  
 QY 54 LeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 73  
 DB 577 GTGACCGACCGACAGTACGCGCATACGAGCGAGCTGAGCAGCGTGGGATCGTGGTG 636  
 QY 74 AspProIle-----AsnThrAsnIysValTrpAla----- 83  
 DB 637 GATACATCGTACAGAAACCGCTGCGATTACGCGCGTCAGCGCATGTCACCAACATCCG 696  
 QY 84 --AlaValGlyMetTyrThrAsnSerTrpAspProAsn----- 95  
 DB 697 GTGACGATCGCGGTGCGAGCAGCATACGACCGCGCTCGCGGTACCGCGGAAGCG 756  
 QY 96 -----AspGlyValIleLeu-----ArgSerAspGln 105  
 DB 757 GGTAGCAGGGTGAGCGTGTATGACGCGCAGCAGCTCTCTGGACGACGACGAGCGATTCG 816  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly----- 119  
 DB 817 TCCGGGAAGTGGACATTTACGCGCACCGC---GGTTGGCGAGCGCGCACTCGATT 873  
 QY 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeu 131  
 DB 874 ACGGTGACTGCGACGGATGCGCGCGCAACGTGAGCACCCTCGAGCGCGCTTCGAGCTG 933  
 QY 132 AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu 151  
 DB 934 AGATCGATACAGCGGCTCCGCGCTGCCCGCGTGAACGCGACGCGGTACGTGCTGCTG 993  
 QY 152 TrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsn-----PhePro 168  
 DB 994 TCGGCACCGCGAAGCGGGCGAGCGGTCAACATCATGATGCGATGCGCGAGCGTACGCG 1053  
 QY 169 AspValGlyTyrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp----- 186  
 DB 1054 GAC---GCCACGTGACGGCTGACCGGCGGCGTCTGGACGTACACGCGCTCGACCCCA 1110  
 QY 187 -----IleGlnGlyValValTrpVal---AlaPheAspLysSerSerSerLeuGly 203  
 DB 1111 CTGCCGATCGGACGGTGTATCGCGGTGACGCGCACGCGCGCGGCGGCAATACGGGCGCG 1170  
 QY 204 GlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSer 223  
 DB 1171 TCTGCTTCGTACCGTACGCGCGGATACGACGCGCAGCGGTGCGCGGTCTATC----- 1224  
 QY 224 ArgAspGlyClyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis 243  
 DB 1225 -----GGTACGTGACGACGATGCG-----GGCTCTGTGTGTGCGATTGCTCT 1272

QY 244 LysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGly 263  
 DB 1273 GCGCGGAGTACTGATGATGCGACCGCACATGTG---AGCGGCACAGCGGAAGCGGCGAGC 1329  
 QY 264 Pro-----TyrAspGlySer-----SerGly 270  
 DB 1330 ACAGTACGCTGTATGACGCGACGACCGTGTGTGGTAGCAGCAGCAGCGCCGTCGGA 1389  
 QY 271 AspValTrpLysPheSerValThrSerGly-----ThrTrpThrArgIleSerProVal 288  
 DB 1390 AAC---TGGACGTTACGCGCACCGCGGTTTGAAGAGCGCGCGACAGCCCTGACCGTC 1446  
 QY 289 ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer-----GlyLeuThrIle 305  
 DB 1447 ACGCCACCGACACGCGCGCAATGTAGTGTTCGAGCAGCGGCTTCGACCTGACGATC 1506  
 QY 306 AspArgGlnHisProAsnThrIleMetVal---AlaThrGlnIleSerTrpTrpProAsp 324  
 DB 1507 GATACGACAGCTCCGTGATTCGACAGTAGTAACCGGACCGAC-----GGT 1551  
 QY 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr 344  
 DB 1552 ACCTCGCTGTCGGCACCGCGGAAGCGCGCAAGCATCAACATCGATACGACGCGGAC 1611  
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364  
 DB 1612 GGCACGCGGATGCCAC-----GTACAGCGCAGCCCGCGCGCGG 1653  
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384  
 DB 1654 TGACGTACACGCGCTCGACCCCGCTCCGCGCT----- 1686  
 QY 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404  
 DB 1687 -----GSCACGCTGATTCGCGTGACG 1707  
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetVal 424  
 DB 1708 GCTACC-----GACGCGCGGCAATACGCGGCCCTCTCTCTCGGTA 1749  
 QY 425 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444  
 DB 1750 -----ACGGTGACGCGGATACGACGCGCGC-----GGTGGCGCGTC 1788  
 QY 445 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464  
 DB 1789 ATCGGTACGTGACGAGCATGCGGCTCTCTGTGTGCGATTCCTCTCGCGCGCAT 1848  
 QY 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
 DB 1849 ACTGATGATCGACCGCGCATTTG---AGCGGCACA-----GCGGACGACGCGC 1893  
 QY 485 ProSerIleValArgAlaGly-----SerPheAspProSer 497  
 DB 1894 AGCAGGTGACGTGTATACGCGCAGCAGCTGTCTCGGCACGACGACGACGCGCATCC 1953  
 QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
 DB 1954 GGA-----AACTGGAGCTTC 1968  
 QY 517 -----GlnGlySerGluProGlyGlyValThr-----Thr 526  
 DB 1969 ACGCCGACTACTGCTGGGTGAGGTGCACACAGCTGACCGTCACCGCAACGATACG 2028  
 QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspPro 546  
 DB 2029 GCGGCAACGTGAGTGTTCGAGCAGCGCTTCGATCTGACGATCGATACGACGCGCTCGG 2088  
 QY 547 GlyGlnProValValTyrAlaVal---GlyPheGlyAsnSerTrpAlaAlaSerGlnGly 565  
 DB 2089 GCGATTCCGACAGTGAACGCGACGCGGTACGTGTTGTCGGGACGCGCGGCGGAT 2148  
 QY 566 ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 585

						2149	GGCAGCGTCAACATGATACGAATGCCGCAGCGCACCCCGGACGCCACCTTAACAGCCGAC	2208
Db								
	::::	:::	:	:	:	586	SerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAia	605
Qy								
	::::	:::	:	:	:	2209	CCCAGCGGCGGTGG-----ACGTACAGCCC--TCGACC	2241
Db								
	:::	:::	:::	:::	:::	606	GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGly	625
Qy								
	:::	:::	:::	:::	:::	2242	CGCTGCCGCTGGTACGGTGATCGCGTGACGGCCACCGATCGCGCAGCAAC-----	2295
Db								
	:::	:::	:::	:::	:::	626	AspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrp	645
Qy								
	:::	:::	:::	:::	:::	2295	-----	2295
Db								
	:::	:::	:::	:::	:::	646	SerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGly	665
Qy								
	:::	:::	:::	:::	:::	2296	-----ACGGGCGCGTCTCGCTCGTAAACGGTGACGGGTGACACCGGACCGGCGGT	2346
Db								
	:::	:::	:::	:::	:::	666	SerSerTyrProAlaValPheValValGlyThr-----lleGlyGlyValThr	681
Qy								
	:::	:::	:::	:::	:::	2347	GGCGCG-----GTCATCGCGACCGTGACCGATGACGTGGGTGTGTGTGT	2391
Db								
	:::	:::	:::	:::	:::	682	GlyAlaTyrArgSer-----AspAspCysGlyThrThrTrpValLeuIleAsn	697
Qy								
	:::	:::	:::	:::	:::	2392	GGTGGATCATTTCCGGCGGCGAGTAGTACCGACCGCCGACATTGAGC-----	2442
Db								
	:::	:::	:::	:::	:::	698	AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu	717
Qy								
	:::	:::	:::	:::	:::	2443	-----GGTACAGCGGAGCGGACGACGACGTGAGC---	2472
Db								
	:::	:::	:::	:::	:::	718	ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAla	737
Qy								
	:::	:::	:::	:::	:::	2473	-----GTGTATGACGGCACGACGTTGTTGGGTACGACGACGCGCCGAC-----	2514
Db								
	:::	:::	:::	:::	:::	738	ProSerGly	740
Qy								
	:::	:::	:::	:::	:::	2515	CCGTCCGA	2523
Db								

## RESULT 11

ACA24001/c  
ID ACA24001 standard; DNA; 5295 BP.

XX  
XX

AC ACA24001;

XX  
DT 10-JUN-2003 (S; not entered)  
XX

4844 GTCACATCGATACGAAATGGCAGCGTACGCCGGAC---GCCACCGTACGGCTGACCCG 4788  
Db  
178 ThrAspThrThrGlyTyrGlnSerAsp-----IleGlnGlyValValTrpVal 193  
Qy  
4787 AGCGCGCTGAGAGCTGACCGCGTGCAGCCCACTCCGATCGGCAGCGTATCGCGTG 4728  
Db  
194 ---AlaPheAspSerSerSerSerLeuGlyGlnAlaSerIleThrIlePheValGly 212  
Qy  
4727 ACGGCCACGAGCGCGCGGCAATACGGCCCGCTCTGCTTCGGTAACTACGGTACGGCGGAT 4668  
Db  
213 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Qy  
4667 AGACGGCACCGCGTCCCGGTCATC-----GGTACGGTACGACGATGCG 4620  
Db  
233 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 252  
Qy  
4619 ---GGTCTGTTGTTGGTGGATGCTCTGGCGGCACTACTGATCGACGCGG 4566  
Db  
253 ValLeuTyrIleAlaThrSerAsnThrGlyGlyPro-----TyrAspGlySer--- 268  
Qy  
4565 ACATTG---AGCGGCACAGCGGAAGCGGCGCAGCGCTCAGCGTGTATGACGGCACGACG 4509  
Db  
269 -----SerGlyAspValTrpLysPheSerValThrSer 279  
Qy  
4508 TTGTTGGGTACGACGACCGCACCGCTCCGGAAC---TGGACGTTACGCGCACCG 4452  
Db  
280 Gly-----ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyr 297  
Qy  
4451 GGTGGGAAGCGGCGCAGACCTGACCTGACCTGACCGCACGACCGCGGCAATGTG 4392  
Db  
298 PheGlyTyrSer-----GlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314  
Qy  
4391 AGTGTTCCGAGCAGCGCTTCGACCTGACGATCAGACGACGCTCGTCCGATCCGACA 4332  
Db  
315 Val---AlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
Qy  
4331 GTGACCGCACCGAC-----GGTACGTCGCTGCGGCGCAGCGGGAAGCG 4287  
Db  
334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Qy  
4286 GCGCAACCATCAACATCATGATAGAAAGCGGCGCAGCGCGGATGCCAC--- 4236  
Db  
354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Qy  
4235 -----GTACACCGACCGCGCGGCGGTGAGCTACACGCGCTCGACCGCGCTG 4185  
Db  
374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Qy  
4184 CCGGCT----- 4179  
Db  
394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTrpAsp 413  
Qy  
4178 -----GGCACGGTGTATGCGTACGGCTACC-----GAC 4149  
Db  
414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Qy  
4148 GCGCGGCAATACCGGCGCGTCTGCTCGGTA-----ACGGTACGGCG 4104  
Db  
434 AspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453  
Qy  
4103 GATACGACGCGCG-----GGTGGCGCGTATCGGTACGTGACGACGATGCGGGC 4050  
Db  
454 PheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 473  
Qy  
4049 TCTGTTGTTGGTGGATGCTCTGCGGCGAGTACTGATGATGCGACGCGCATTTG--- 3993  
Db  
474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValargAlaGly--- 492  
Qy  
3992 AGCGGCACA-----GCGGAAGCAGGCGACGCGTGGAGCGGTATGACGCGCAGC 3945  
Db  
493 -----SerPheAspProSerSerGlnProAsnAspArgHisValAla 506  
Qy  
3944 ACGTGCTCGGCACGACGACGCGGACCATCCGGA----- 3909  
Db

507 PheSerThrAspGlyGlyLysAsnTrpPhe-----GlnGly 518  
Qy  
3908 -----AATGACGTTACGCCGACTACTGCACTGGTGAGGTT 3870  
Db  
519 SerGluProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSerAlaAsp 535  
Qy  
3869 GCACACAGCTGACCGTCAACGCAACGATACCGCGGCAACGTGAGTGTTCGAGCAGC 3810  
Db  
536 GlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal--- 554  
Qy  
3809 GCCTTCGATCTGACGATCGATACGACGGCTCCGCGGATCCGACAGTGAACGCGCGAC 3750  
Db  
555 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 574  
Qy  
3749 GGTACGTGTTGTCGGCACCGCGCAAGCGGTCGACGGTCAACATCGATACGAATGCG 3690  
Db  
575 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 594  
Qy  
3689 GACGGCACGCGGACGCGCACCGTAACACGCGGACCCGACGCGCGCTGG--- 3642  
Db  
595 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 614  
Qy  
3641 -----ACGTACACGCCC---TCGACCCCGCTGCCGCTGGTACGCGTATCGCG 3597  
Db  
615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 634  
Qy  
3596 GTGACGGCACCGATCGGCGACGCAAC----- 3570  
Db  
635 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 654  
Qy  
3569 -----ACGGGCGCGTCTGCTCTGCTCTG 3552  
Db  
655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 674  
Qy  
3551 GTAACGCTGACGGGTGACACGCGGACCGCGGTCGTTGTTGGTGGATCATTCGCGGCGAGTACT 3447  
Db  
675 GlyThr-----IleGlyGlyValThrGlyAlaTyrArgSer--- 686  
Qy  
3506 GGCACCGTACGATGACGTGGGTCTGTTGTTGGTGGATCATTCGCGGCGAGTACT 3447  
Db  
687 AspAspGlyGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706  
Qy  
3446 GATGACGCGACGCGACATTGAGC----- 3423  
Db  
707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726  
Qy  
3422 GGTACGCGGAGGACGACGACGACGATGAGC-----GTGTATGACGCGACGACGAGTTG 3372  
Db  
727 ArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Qy  
3371 TTGGTACGACGACGCGCGAC-----CCGTCCGGA 3342  
Db

## RESULT 12

ACA40315  
ID ACA40315 standard; DNA; 9903 BP.  
XX  
AC ACA40315;  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX  
DE Prokaryotic essential gene #21972.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Mycobacterium tuberculosis.  
OS  
XX  
FN W020027183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI: 2003-029926/02.  
 DR P-PSDB; ABUS6445.  
 DR  
 DR  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 14; SEQ ID NO 28185; 1766pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 9903 BP; 2129 A; 3380 C; 2727 G; 1667 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 0.018 Length: 9903  
 Score: 192.00 Matches: 194  
 Percent Similarity: 35.04% Conservative: 94  
 Best Local Similarity: 23.60% Mismatches: 323  
 Query Match: 4.76% Indels: 213  
 DB: 7 Gaps: 43  
 US-09-917-376-3 (1-740) x ACM40315 (1-9903)  
 Qy 10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly----- 27  
 Db 7399 AAGCGAACAATCGCGCGCGCAACATCGCGAC-----TTTAACGTCGGATCGCA 7449  
 Qy 28 -----AlaProGlyIleLeuTyrValArgThrAspIleGly-----GlyMet 41  
 Db 7450 AACACCGGTCGCGGGCTAACCGGGCTGTGTCAACAACATCGGTATCGGCAACACCGGCAAC 7509

Qy 42 TyrArgTrpAspAlaAlaAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 Db 7510 TACAACATCGGTGCGCAACACCGTAACACAC-----ATCGGCTTC 7554  
 Qy 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp----- 74  
 Db 7555 GGCAACACCGGCAACACACATCGGCATCGGCTGTCCGGCGACACACAGATCGGGTTC 7614  
 Qy 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrTrpAsnSerTrpAsp 93  
 Db 7615 GGCCCGCTGAACCGCGG-----ATCGCAACATCGGCTGTTC---AACCTGGCGGAC 7665  
 Qy 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113  
 Db 7666 AACAACTTTGCG----- 7677  
 Qy 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133  
 Db 7678 -----ATGGCCAAACCGGCGCAACTTCAACACGAGGCGATGCGC 7713  
 Qy 134 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp 152  
 Db 7714 AACACCGGCAACACACATCGGCTGTTCACACCGGCAACACACACGTCGCGATCTCG 7773  
 Qy 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172  
 Db 7774 CTGACCGGCGACCGGCTTCCGGCTTCACTCCCTGAACCTCGCGCGCGCAACACCGGT 7833  
 Qy 173 TyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192  
 Db 7834 TTCTTCAACTCGGCGACCGCGCAACACCGCG----- 7863  
 Qy 193 ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211  
 Db 7864 ---TTGTTCACCTCGGCGACCGCGCAACACCGCTGTTCACCTCGGCGCGCAACGTC 7920  
 Qy 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
 Db 7921 GGCAACACCGGCGCGCGCGCGGCTTCCGCGCTCGGCGCTATCGGCGACCGAGGTG 7980  
 Qy 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251  
 Db 7981 GGCAATCGGCGCGCAACCACTCGGCGAGTTTC---AACATCGGCTGTTCACCTCGGCGAC 8037  
 Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGly-As 271  
 Db 8038 GGCAATCGGCGCGCAACCACTCGGCGACCGCGCAACGTCGCGCATCGGCAACACCGGAC 8097  
 Qy 271 pValTrpLysPheSerValThrSerGlyThr----- 282  
 Db 8098 GGCAACACCGGCGATCGGAACACGCGGCAACTACAACACCGGCTTCTCAACCGGCGCTG 8157  
 Qy 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296  
 Db 8158 GTCAACACCGGCGATCGGCGCAACCGGCGCAACCAACACCGGCGCTTCAACATCGGCGAC 8217  
 Qy 296 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316  
 Db 8218 TTCAACACCGGCGATCGGCGCAACCGGCGCGCTTCAACACCGGCTTCTCAACACCGGTAGC 8277  
 Qy 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331  
 Db 8278 TACAACACCGGCGATCGGCGCAACCGGCGAGACTACGCGCGCGCTTCTATCACCGGCGAC 8337  
 Qy 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpTrpSer--- 345  
 Db 8338 ATGAACAACCGCTTGTCTCTGCGCGCGCGACCGCGAGGCGCTGCTGCGCGCGCAACTACAC 8397  
 Qy 346 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaG 360  
 Db 8398 ATCAACATCGAGCGACCTGCGCGCTTCTCAATGTTCGACATCCCGGTCACACATCCCATC 8457  
 Qy 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380

8458 ACCG---GGACATACCAATGCTCCATCCCGCCATTACGTTCCCCAGA-ATC-----8508  
QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 399  
Db 8509 -----GACGCCAGCGGAAGGTGCACATAGGCATCTCAGTGGCACCGTCTTGGCCCGGT 8564  
QY 399 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418  
Db 8565 CGGTCCGATCACCTTCGATGGCGGAGCGCTGCGGCCCGCTGGACACACCCATCGAAAT 8624  
QY 418 eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuSerPr 438  
Db 8625 TGACTTCGGCCCC-----TCGCGCGCGATCAACCTCAACATCGGCAA 8666  
QY 438 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs 458  
Db 8667 GCCCGAGGCTCCACCGTATCAACATCGTGGCGCGCGCGC-----8709  
QY 458 pValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVa 478  
Db 8710 -----GCCGCGCCGATC-----AGCAT 8726  
QY 478 lAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 498  
Db 8727 TCGCATCATCGACTTGGCGCCAGCG-----CCCGGCTTCTTCAACGCCACAC 8774  
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnI 518  
Db 8775 CGGCCCG-----TCGTGGCGCTTCTCAACTGGGGTCTGG 8810  
QY 518 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValAlaAl 532  
Db 8811 CACGCATCGGCTTCTGCTGAATCTGGCAACAACTCGGGCTTACACTTCGCCACTAG 8870  
QY 532 aserAlaAspGlySerArgPheVal-----TrpAlaProGl 544  
Db 8871 CACATGGGAATTCGGGCTTCAAACTATGGTTCGCTGCAGTGGGCTGGCG-----8925  
QY 544 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 560  
Db 8926 -----AATTTGGCAACAGCATCTCGGGCATCTA 8954  
QY 560 pAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 580  
Db 8955 CAACACCGGCTGGGAGCACCGGCAATGTC-----TCGGGCTTGTCTCAACATCGG 9005  
QY 580 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVa 598  
Db 9006 CACCAACCTGGTGGTGGTTCGACAGACGCCCG-----9039  
QY 598 lThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPhe- 617  
Db 9040 -ACCGAGACGACCTTCAGCGTGGGCTTGGCAACCTCGGGTTCT---GGAATCTGGGTAG 9095  
QY 618 -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 9096 CGCAACATCGCAACTACAACTGGGCGACGCCAACATCGGCGTCTACAACTCGGCAG 9155  
QY 630 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerA 647  
Db 9156 CGCAACATCGCGGACTTCAACTGGGCGAGCGCAACATCGGCGACTTCAACTGGGCAG 9215  
QY 647 lalleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerS 667  
Db 9216 CGCCAACA-----TCGGCAGCTCC-AAACATCGGTTGGCAACGTCGGTCCGGG---9264  
QY 667 erTyrProAlaValPheValGlyThrIleGly---GlyValThrGlyAlaTyrArgS 686  
Db 9265 -----CTGACGGCGGCGCATCGGCAACATCGGCTTCGGCAACCGGAAACGGAACA 9316  
QY 686 erAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 706

Db 9317 TCGGCATCGCAATACCGGCACCGGCAACATCGGCTTCGGCAACACCGGAACGAAACA 9376  
QY 706 rpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgValTyrIle-GlyThrA 725  
Db 9377 TCGGCATCGGCTGACCGGCGACACCATGACCGGGTTTCGGGCTTGAACATCGGCACCG 9436  
QY 725 snGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 9437 GCAACATCGGCTATTCAACTCGGCACCGGCAACATCGGCTTCGGCAACTCCGCGC 9492

## RESULT 13

AAI99682\_04/c  
Continuation (5 of 45) of AAI99682 from base 400001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

## Alignment Scores:

Pred. No.:	0.285	Length:	110000
Score:	192.00	Matches:	194
Percent Similarity:	35.04%	Conservative:	94
Best Local Similarity:	23.60%	Mismatches:	323
Query Match:	4.76%	Indels:	213
DB:	4	Gaps:	43

US-09-917-376-3 (1-740) x AAI99682\_04 (1-110000)

QY 10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly-----27  
|||||

Db 27279 AACCGCAACATCGCGCGCAACATCGCGAC-----TTTAACGTGGGATCGCA 27229  
Qy 28 -----AlaProGlyLeuTyrValArgThrAspIleGly-----GlyMet 41  
Db 27228 AACACCGGTCCGGGGTCAACGGCGGTGTCAACAACTCGTATCGGCAACACCGGCAAC 27169  
Qy 42 TyrArgTrpAspAlaAAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 27168 TACAACATCGGTGTGGCAACACCGGTAACCAAC-----ATCGGCTTC 27124  
Qy 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp----- 74  
Db 27123 GGCAACACCGGCAACAAACATCGCATCGGCTGTCCGGGCAACACAGATCGGGTTC 27064  
Qy 75 ---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 93  
Db 27063 GGCCCGCTGAACCGCGC-----ATCGCAACATGGCGCTGTC---AACCTGGGCGAC 27013  
Qy 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113  
Db 27012 AACAACTTTGGC----- 27001  
Qy 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133  
Db 27000 -----ATGCCAACCGCGGCAACTTCAACAGGGCATTTGCC 26965  
Qy 134 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp 152  
Db 26964 AACACCGGCAACAAACATCGGCTGTTCACACCGGCAACAAACAGTCGGCATCTGG 26905  
Qy 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172  
Db 26904 CTGACCGGCGACCGCTTGTCCGGCTTCAGCTCCCTGAATCTCCGGCGCGCAACACCGGT 26845  
Qy 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192  
Db 26844 TTCTTCAACTCCGGCAGCGCAACACCGGC----- 26815  
Qy 193 ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211  
Db 26814 ---TTGTTCAACTCCGGCAGCGCAACACCGGCTTGTTCACCTCGGCGCAACAGTC 26758  
Qy 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
Db 26757 GGCAATCGGCAACATGGCGCGGCTTCGGGCTTCGGGCTATCCGGCGACAGCAGGTG 26698  
Qy 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251  
Db 26697 GGCAATCGGCGCACCAACTCGGCGAGTTTC---AACATCGGCTGTTTAACTCGGCGACC 26641  
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGly-As 271  
Db 26640 GGCAATGTGGCATCGGCAACTCGGCGAGCGCAACCTCGGCATCGGCAACACCGGCAAC 26581  
Qy 271 pValTrpLysPheSerValThrSerGlyThr-----Tr 282  
Db 26580 GGCAACACCGGCATCGGAAACAGCGGCAACTACAACACCGGCTTGTCTCAACCGGCGCTG 26521  
Qy 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296  
Db 26520 GTCAACACCGGCATCGCAACCGGCGCAACCAACACCGGCTGTTCACATCGGCGACC 26461  
Qy 296 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316  
Db 26460 TTCAACACCGGCATCGGCAACCGGCGCACTACAACACCGGCTCTTCAACACCGGTAGC 26401  
Qy 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331  
Db 26400 TACAACACCGGCATCGCAACCGGCGAGACTACGGCACCGGCGGTTCATCACCGGCGAGC 26341  
Qy 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer-- 345  
Db 26340 ATGAACAAACGGCTTGTCTTGGCGCGCGACCGGCGGCGCTGTGGCGGCAACTACACC 26281

Qy 346 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGl 360  
Db 26280 ATCAACATCGAGCACCTGCGCGGTCTCTCAATGTGACATCCCGTCAACATCCCATC 26221  
Qy 360 uproTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380  
Db 26220 ACCG---GCGACATCAACCAATGTCTCCATCCCGCATTTACGTTCGCCAGA-ATC----- 26170  
Qy 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 399  
Db 26169 ---GAGCCAGCGGAGCGTCGACATAGGCATCTCAGTCGACCGGTTCGCCCGGT 26114  
Qy 399 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI 418  
Db 26113 CGGTCCGATCACCTGATGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26054  
Qy 418 eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPr 438  
Db 26053 TGACTTCGGCGCC-----TCGCGCGCGGATCAACCTCAACATCGGCA 26012  
Qy 438 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs 458  
Db 26011 GCCGACGGCTCCACCGGTGATCAACATCGTGGCGCGCGCGCGCGCGCGCGCG 25969  
Qy 458 pValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVa 478  
Db 25968 -----GCCGCGCGCATC-----AGCAT 25952  
Qy 478 lAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 25951 TCCGATCATCGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25904  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnG 518  
Db 25903 CGGCGCG-----TCGTGGCGCTTCTCACTGGGTGCTGG 25868  
Qy 518 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValAlaAl 532  
Db 25867 CAGCGCATCGGCTTGTCTGAATTCGCGCAACAACTCGGCGCTTACAACTTCGCCACTAG 25808  
Qy 532 aSerAlaAspGlySerArgPheVal-----TrpAlaProGl 544  
Db 25807 CAGCATGGGAATTCGGGCTTCCAAAATATGGGTCTGCTGCACTCGGCGTGGCGCG 25753  
Qy 544 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 560  
Db 25752 -----AATTGGGCAACAGCATCTCGGCGCATCTA 25724  
Qy 560 pAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 580  
Db 25723 CAACACCGGCTTGGGAGCACCGGCAATGTC-----TCGGGTCTTCAACATCGG 25673  
Qy 580 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVa 598  
Db 25672 CACCAACCTGGCTGGGTGGTTCAGAACCGGCGCG----- 25639  
Qy 598 lThrPheGlnProValAlaAlaGlyLeu- ProSerSerGlyAlaValGlyValMetPhe- 617  
Db 25638 -ACCGAGACGACCTTCAGCGGTGGGCTTGGCGCAACCTCGGCTTCT---GGAAATCTGGGTAG 25583  
Qy 618 -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 25582 CGCAACATCGGCAACTACAACTGGGCGAGCGGCAACATCGGCGCTTACAACTGGGCGAG 25523  
Qy 630 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerTrpSerA 647  
Db 25522 CGCCAAACATCGGCGACTTCAACCTGGGCGAGCGGCAACATCGGCGACTTCAACCTGGGCGAG 25463  
Qy 647 lalleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerS 667  
Db 25462 CGCCAAACA-----TCGCGAGCTCC- AACATCGGCTTGGCAACGTCGCTCGGCGG- 25414

Qy	667	erTyrProAlaValPheValValGlyThrIleGly---GlyValThrGlyAlaTyrArgS	686
Db	25413	-----CTGACGGCGGCATCGGCACATCTGGCTTCGGCAACACCGGAACACGAAACA	25362
Qy	686	erAspAspCysGlyThrThrTpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT	706
Db	25361	TCGGCATCGGCATATACCGGCACCGGCACATCTGGCTTCGGCAACACCGGAACACGGAACA	25302
Qy	706	rpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgValTyrIle-GlyThrA	725
Db	25301	TCGGCATCGGCTGACCGCGCGACCATCATGACCGGGTTTGGCGGGCTGGAACTCGGGCACCG	25242
Qy	725	snGlyArgGlyIleVal-----TyrGlyAspIleGlycIcYalaProSerGly	740
Db	25241	GCAACATCGGGCTATTCAACTCCGGCACCGGCAACATCGGCTTCGGCAACTCCGCG	25186

## RESULT 14

ACA38418  
ID ACA38418 standard; DNA; 3324 BP.

XX  
AC ACA38418;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #20075.

Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW  
drug design; gene.  
KW

OS *Mycobacterium bovis*.

PN WO200277183-A2.

03-OCT-2002.

21-MAR-2002: 2002WO-US009107.

PR 21-MAR-2001: 2001US-00815242.

PR 25-OCT-2001: 2001US-0342923P.  
PR 08-SEP-2001: 2001US-0094853P.

PK 08-FEB-2002; 2002US-00072831:  
PB 06-MAR-2002; 2002US-0362699P:

PA (ELIT-) ELITRA PHARM INC.

Wang L., Zamudio C., Malone C.

FI wall D, Frawick JD, Car  
XX

DR WPI; 2003-029926/02.  
DR P-PSDB: ABU34548.

New antisense nucleic acids

PT isolate homologous HU  
PT isolate candidate

PS Claim 14: SEO ID NO 26288: 1766pp: English: XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: the sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC [www.int/pub/abstract/pct\\_sequences](http://www.int/pub/abstract/pct_sequences)

Sequence 3324 BP: 726 A: 1184 C: 886 G: 528 T: 0 U: 0 Other:

Alignment Scores:		
Pred. No.:	0.0169	Length: 3324
Score:	184.00	Matches: 205
Percent Similarity:	35.53%	Conservative: 89
Best Local Similarity:	24.81%	Mismatches: 313
Query Match:	4.56%	Indels: 221
DB:	7	Gaps: 48

US-09-917-376-3 (1-740) x ACA38418 (1-3324)

Qy	10	AsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGly-----	27
Db	181	AACCGCAACATCTCGCGCGCAAAACATCGGCAC-----TTTAACGTGCGATCGCA	231
Qy	28	-----AlaProGlyIleLeuTy=ValArgThrAspIleGly-----GlyMet	41
Db	232	AACACCGTCTCGGGGTAAACGGCGGTGTCAACAACATCGGTATCGGCAACACCGCAAC	291
Qy	42	TyrArgTrpAspAlaIaAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp	60
Db	292	TACAACATCGGTCTCGGCAACACCGGTAACTACAAC-----ATCGCTTTC	336
Qy	61	AsnAsnTrpGlyTy=AsnGlyVal---ValSerIleAlaAlaAsp-----	74
Db	337	GGCAACACCGGCACAAACAACTCGGCATCGCCTCCGCGCAACACAGATCGGGTTC	396
Qy	75	---ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTy=ThrAsnSerTrpAsp	93
Db	397	GGCCCGCTGAACGCGGC-----ATCGCAACATGGGCTGTTC---AACCTGGCGCAC	447
Qy	94	ProAsnAspGly---AlaIleLeuArgSerSerAspGlnGlyAla-----ThrTrpGln	110
Db	448	AACAACATTTGGCATGGCCACACGCGGCAACTTCAACAGGGCATTTGCCAACACCGGCAAC	507
Qy	111	IleThrProLeuProPheLeuLeuGlyGlyAsnMetProGlyTy=ArgGlyMetGlyGlu=Arg	130
Db	508	AACAACATCGCTGTCTTCAACACCGGCACAAACAACGTCTGGCATCGGCTG-----	558
Qy	131	LeuAlaValAspPro=AsnAsnAspAsnIleLeuTy=TrpPheGlyAlaProSerGlyLysGly	150
Db	559	-----ACCGCGCACGCG-----ACCGCGCACGCG	570
Qy	151	Leu-----TrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePro	168
Db	571	TTGTCCGGCTTCAGCTCCTCGTAATCTCGGCACC-----GGCAACACCGGTTCCTTC	621
Qy	169	AspValGlyThrTy=IleAlaAsnProThrAspThrThrGlyTy=TrpGlnSerAspIleGln	188
Db	622	AACCTCGGCACC-----GCCAAC-----ACCGCG-----	645
Qy	189	GlyValValTrpValAlaPheAspLysSerSerSerSerLeuLeuGlyGln---AlaSerLys	207
Db	646	-----TTGTTCATCTCGGCACCGGCAACACCGGCTTGTTCACCTCGGGC	690
Qy	208	ThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly	227



Db 691 ACCGGAAACGTCGGCATCGGCAACATCGGCAACCGCGGCTTCGGCGTCGGCTATCCGGC 750  
Qy 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPhe 247  
Db 751 GACAGCCAGGTGGGCATCGGCGGACCACTCGGCGAGTTTC---AACATCGGCTTGT 807  
Qy 248 AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267  
Db 808 AACTCGGCGCACCGGCAATGTCGGCATCGGCAACTCGGCGACCGGCAACGTCGGCATCGGC 867  
Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThr----- 281  
Db 868 AACACCGGACCGGCAACACCGGCATCGGAACAGCGCAACTCAACACCGGCTTGTCTC 927  
Qy 282 -----Trp-----ThrArgIleSerProValProSerThrAs 292  
Db 928 AACCGGGGCTGTGTCAACACCGGCATCGCCNACCGGGCAACCAACACCGGCTGTTC 987  
Qy 292 pThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnTh 312  
Db 988 AACATCGGCACCTTCAACACCGGCATCGCCAAACCGGGCCACTACAAACACCGGCTCTAC 1047  
Qy 312 rIleMetValAlaThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSe 330  
Db 1048 AACACCGGTAGTACAAACACCGGCATCGGCAACCGCGAGACTACGGCACCGGCGCTTC 1107  
Qy 330 rThr-----AspGlyGlyAlaThrTrpThrArgIleTrpAs 342  
Db 1108 ATACCGCGCAGCATGAACACCGGCTTCTCTGGCGCGCGGACCGCGGCGCTCTGGCG 1167  
Qy 342 pTrpThrSer-----TyrProAsnArgSerLeuArgTyrValLeuAs 356  
Db 1168 GCCAACTACCACTACCATCACCATCGAGCGACTCGCGCTTCTCAATGTGACATCCCGGTC 1227  
Qy 356 pIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerPr 376  
Db 1228 AACATCCCCATCACCG---CGGACATCAACCAATCTCTCCATCCCGCGCATTACGTTCCCC 1284  
Qy 376 oLysLeuGlyTrpMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMe 395  
Db 1285 AGA-ATC-----GACCGCCGGAAGCGTCGATAGCATTCCTCAGTCGACCGCT 1334  
Qy 395 tLeuTyrGlyThrGly---AlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSe 414  
Db 1335 CTGGGCCCCGTCGTCGATCACCTCGATGGCGGGGACCGTCGCGCCCGCTGGACAC 1394  
Qy 414 rGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAs 434  
Db 1395 ACCCATCGAAATTGACTTCGGCCCC-----TCGCGCGCGATCAACCT 1436  
Qy 434 pLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPh 454  
Db 1437 CAACATCGGCAAGCCGCGCTCCACCTGATCAACATCGTGGCGCGCGCGC----- 1491  
Qy 454 eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474  
Db 1492 -----GCCGCGCCGATC----- 1503  
Qy 474 rGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPh 494  
Db 1504 -----AGCATTCGATCATCGACTTCGGCGCACG-----CCGCGCTTCTT 1544  
Qy 494 eAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAs 514  
Db 1545 CAACGCCACACCGGCGCG-----TCGTCGGGCTTCTCCTCAA 1580  
Qy 514 nTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr----- 529  
Db 1581 CTGGGTGCTGGAGCGCATCGGCTTCTGAACTTCGGCAACAACTCGGCGCTCTACAA 1640  
Qy 530 -----ValAlaAserAlaAspGlySerArgPheVal----- 540

Db 1641 CTTGCCCACTAGCAGCATGGGAAATTCGGGCTTCAAAACATATGGTCTCGCTAGTCGGG 1700  
Qy 541 -TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-- 559  
Db 1701 CTGGGCG-----AATTTGGGCAACAGCAT 1724  
Qy 560 -----TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspAr 576  
Db 1725 CTCGGGCACTTACAAACACCGCTTGGGAGCACCGGCAAAATGTC-----TCGGGCTT 1775  
Qy 576 gValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerTh 594  
Db 1776 GCTCAACATCGCGCAACACTGGCTGGTGTGCAGAACCGGCCG----- 1821  
Qy 594 rAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValG 614  
Db 1822 -----ACCGAGACGACCTTCAGGTGGCTTGGCAACCTCGGTTCT---G 1865  
Qy 614 lValMetPhe-----HisAlaValProGlyLysGluGlyA 626  
Db 1866 GAATCTGGGTAGCGCAACATCGGCAACTACAACTGGGCGCGGCAACATCGGCTCTA 1925  
Qy 626 spLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyLys 643  
Db 1926 CAACCTGGCGAGCGCAACATCGGCGACTTCAACCTGGGCGAGCGCAACATCGGCGACTT 1985  
Qy 643 erSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerA 663  
Db 1986 CAACCTGGCGAGCGCAACA-----TCGGAGCTCC-AACATCGGTTTCGGCAACGTCG 2038  
Qy 663 laProGlySerSerTyrProAlaValPheValValGlyThrIleGly---GlyValThrG 682  
Db 2039 GTCCGGGG-----CTGACGGCGGCATCGGCAACATCGCTTCGGCAACACCG 2086  
Qy 682 lValAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisG 702  
Db 2087 GAAACGGAACATCGGCATCGGCAATACCGGCAACATCGGCTTCGGCAACACCG 2146  
Qy 702 lntTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgValTy 721  
Db 2147 GAAACGGAACATCGGCATCGGCTGCGGCGGACCAACATGACCGGTTTCGGCGGTGA 2206  
Qy 721 rIle-GlyThrAsnGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlap 738  
Db 2207 ACTCGGCGACCGCAACATCGGCTATTCAACTCGGCGACCGCAACATCGGCTTCGGCA 2266  
Qy 738 roSerGly 740  
Db 2267 ACTCGGC 2274

## RESULT 15

AAI99683\_04/c

Continuation (5 of 44) of AAI99683 from base 400001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000
WP	AAI99683_08	800001	910000
WP	AAI99683_09	900001	1010000
WP	AAI99683_10	1000001	1110000
WP	AAI99683_11	1100001	1210000
WP	AAI99683_12	1200001	1310000
WP	AAI99683_13	1300001	1410000
WP	AAI99683_14	1400001	1510000
WP	AAI99683_15	1500001	1610000
WP	AAI99683_16	1600001	1710000
WP	AAI99683_17	1700001	1810000



```
Db 26073 CAAGCCACACCGCGCGG-----TCGTCGGGCTTCCTCAA 26038
Qy 514 nTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThr----- 529
Db 26037 CTGGGGTGCTGGCAGCGCATCGGGCTTCTGAACCTCGGCAACAACCTCGGGCTCTACAA 25978
Qy 530 -----ValAlaAlaSerAlaAspGlySerArgPheVal----- 540
Db 25977 CTTGGCCACTAGCAGCATGGGAAATTCGGGCTTCAAAACTATGGGTCTGCAGTCGGG 25918
Qy 541 -TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-- 559
Db 25917 CTGGGCG-----AATTGGGCAACAGCAT 25894
Qy 560 -----TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspAr 576
Db 25893 CTCGGGGCATCTACACACCGGCTTGGGAGCAGCAGCGCAATGTC-----TCGGGCTT 25843
Qy 576 gValAsnProGlyThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerTh 594
Db 25842 GCTCAACATCGGCACCAACCTGGCTGGTGTGTTGCAGAACGGCCG----- 25797
Qy 594 rAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValG 614
Db 25796 -----ACCGAGAGACCTTCAGCGTGGGCTTGGCCAACTCGGGTTCT--G 25753
Qy 614 lyValMetPhe-----HisAlaValProGlyLyGluGlyA 626
Db 25752 GAATCTGGGTAGCGCAACATCGGCAACTACACCTGGGCGCGCCMAACATCGGCGTCTA 25693
Qy 626 spLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlyS 643
Db 25692 CAACCTGGGCGCGCCAAACATCGCGACTTCAACCTGGGCGCGCCAAACATCGCGACTT 25633
Qy 643 erSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerA 663
Db 25632 CAACCTGGGCGCGCCAAACA-----TCGGCAGCTCC-AAACATCGGGTTCGGCAACGTCG 25580
Qy 663 laProGlySerSerTyrProAlaValPheValValGlyThrIleGly---GlyValThrG 682
Db 25579 GTCGGGG-----CTACGGCGCGCATCGGCAACATCGGCTTCGGCAACACCG 25532
Qy 682 lyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisG 702
Db 25531 GAAACGAAACATCGGCATCGGCAATACCGGCACCGCAACATCGGCTTCGGCAACACCG 25472
Qy 702 lnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp--HisAlaAsnLeuArgArgValTy 721
Db 25471 GAAACGAAACATCGGCATCGGCTGACCGCGCAGACCATGACCGGGTTCGGCGGTGGA 25412
Qy 721 rIle-GlyThrAsnGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlap 738
Db 25411 ACTCGGCGACCGCAACATCGGGCTATTCAACTCGGCGACCGGCAACATCGGCTTCGGCA 25352
Qy 738 roSerGly 740
Db 25351 ACTCGGCG 25344
```

Search completed: May 12, 2004, 05:39:13  
Job time : 1279.48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 6950.51 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWTSNVAIGGGGFD.....YIGTNGRGIYVGDIGGAPSG 740

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_epool/US0917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0917376@cgn\_1\_1\_5334@runat\_11052004\_114535\_28249 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_gssI:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	615.5	15.3	704	14	CF872104	CF872104 trico29xo
2	615.5	15.3	761	14	CB902332	CB902332 trico29xo
3	587.5	14.6	747	14	CF876916	CF876916 trico074xd
4	587.5	14.6	814	14	CB905388	CB905388 trico074xd
5	415.5	10.3	707	14	CF880713	CF880713 trico082xn
6	415.5	10.3	782	14	CB907625	CB907625 trico082xn
7	385.5	9.6	693	14	CF882065	CF882065 trico29xo
8	361.5	9.0	751	14	CF868882	CF868882 trico16xm
9	361.5	9.0	803	14	CB898982	CB898982 trico16xm
10	338	8.4	738	14	CF866388	CF866388 trico06xe
11	338	8.4	794	14	CB896354	CB896354 trico06xe
12	314.5	7.8	546	10	BF072664	BF072664 NCSM3H773
13	306.5	7.6	389	13	BU639045	BU639045 mgcw011xd
14	286.5	7.1	713	14	CF875815	CF875815 trico039xk
15	286.5	7.1	782	14	CB904767	CB904767 trico039xk
16	282.5	7.0	719	14	CF881775	CF881775 trico085xf
17	282.5	7.0	775	14	CB908435	CB908435 trico085xf
18	278.5	6.9	929	14	CF885920	CF885920 trico085xf
19	222	5.5	796	14	CF391910	CF391910 RTDRJ_10
20	183	4.5	3057	11	AK044947	AK044947 Mus muscu
21	180	4.5	4355	29	AY416870	AY416870 Mus muscu
22	171.5	4.2	296	13	BU641770	BU641770 mgmk008xa
23	170	4.2	4631	11	AK082944	AK082944 Mus muscu
24	166.5	4.1	338	9	AW064456	AW064456 SPI055 KR
25	153.5	3.8	3623	11	AK083848	AK083848 Mus muscu
26	153.5	3.8	3751	11	AK080277	AK080277 Mus muscu
c 27	153.5	3.8	4101	29	AY402312	AY402312 Homo sapi
28	151	3.7	4654	11	AK086725	AK086725 Mus muscu
c 29	148.5	3.7	1178	28	CC203127	CC203127 CH261-8B4
30	140	3.5	3782	11	AK041860	AK041860 Mus muscu
c 31	139.5	3.5	5367	11	AK084803	AK084803 Mus muscu
c 32	138	3.4	5073	29	AY407742	AY407742 Homo sapi
33	137.5	3.4	1886	11	AK009469	AK009469 Mus muscu
34	137	3.4	1581	11	BC029959	BC029959 Homo sapi
35	136.5	3.4	1569	29	AY407967	AY407967 Homo sapi
36	136	3.4	1674	29	AY412937	AY412937 Mus muscu
37	134	3.3	426	13	BQ751758	BQ751758 EST632321
38	133.5	3.3	2545	14	CD014011	CD014011 90138962
39	132	3.3	2565	11	AK030673	AK030673 Mus muscu
40	132	3.3	3264	29	AY404484	AY404484 Homo sapi
41	132	3.3	4611	29	AY418194	AY418194 Homo sapi
42	131.5	3.3	3762	11	BC039264	BC039264 Homo sapi
43	131.5	3.3	4132	11	AK032858	AK032858 Mus muscu
c 44	131.5	3.3	6914	11	BC042295	BC042295 Homo sapi
45	130	3.2	1723	11	AK089973	AK089973 Mus muscu

ALIGNMENTS

RESULT 1  
CF872104  
LOCUS CF872104 704 bp mRNA linear EST 31-OCT-2003  
DEFINITION trico29xo13.b1 T.reesei mycelial culture, Version 6 October 2003  
ACCESSION CF872104  
VERSION CF872104.1 GI:38126786  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
REFERENCE 1 (bases 1 to 704)

AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D., Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and Dean,R.A.  
 TITLE Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu  
 Seq primer: LT-F1 primer.  
 Location/Qualifiers  
 1..704  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="tric029xo13"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,4e-43 Length: 704  
 Score: 615.50 Matches: 118  
 Percent Similarity: 75.61% Conservative: 37  
 Best Local Similarity: 57.56% Mismatches: 43  
 Query Match: 15.25% Indels: 9  
 DB: 14 Gaps: 4

US-09-917-376-3 (1-740) x CF872104 (1-704)  
 QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 DB TTTTCATCGAAGACAGTCAAGCTCGCGCGCGCGCGCTTCGTCCCGGCATCATCTTC 157  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43  
 DB CATCCCAAGACAAAGGGGTAGCATATGCACGAACAGATATGGGGCTGTACCGCTC 217  
 QY 44 TrpAspAlaAlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGlyTrpAsn 61  
 DB TTCTCCGCGACTCATGGACCGCC-GTCACGGATGGGATTGCTGATAATGCCGCTGGCAC 276  
 QY 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleVal 81  
 DB AACATGG-----GGCATCGCTGTTCGCTGTTCGCGAGGAGCATCAAAAGGTG 327  
 QY 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101  
 DB TATCGCGAGTCGCACTATACGACAGCTGGGATCCGAGTATGAGGACATCATTCGC 387  
 QY 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsn 121  
 DB TCGTCAGACCGCGCGCAACGTGTCCTTCCACCACTTGCCTTCAAAGTCGGGGTAAAC 447  
 QY 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspIleLeu 141  
 DB ATGCAGACCGCGAGCGGAGAGCGTCTGGCTGTTCGATCCGGCCAACTCCAAACATCATC 507  
 QY 142 TyrPheGlyAlaProSerGlyIleValLeuTrpArgSerThrAspSerGlyAlaThrTrp 161  
 DB TACTTTGTGTCCTCAGGAACCGGCTCTGGAAGTCTACGAGCGCGGTGACCTTT 567  
 QY 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181

Db 568 TCCAAGGCTCTCGTTCACGGCAACTGGAGCTACATCCACGCCGAGTGATTCACAC 627  
 QY 182 GlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSerSer 201  
 Db 628 GGCTACCAAGGATCAGCAAGGACTCATGTGGTTACGTTCCACTC-ACCAGCAGCAG 686  
 QY 202 LeuGlyGlnAlaSer 206  
 Db 687 ACCGGNGGAGCCACG 701

RESULT 2  
 CB902332 761 bp mRNA linear EST 02-JUL-2003  
 LOCUS tric029xo13 T.reesei mycelial culture, Version 3 april Hypocrea  
 DEFINITION jecorina cDNA clone tric029xo13, mRNA sequence.  
 ACCESSION CB902332  
 VERSION CB902332.1 GI:30116990  
 KEYWORDS EST.  
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
 ORGANISM Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 REFERENCE 1 (bases 1 to 761)  
 AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,  
 Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,  
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,  
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.  
 TITLE Transcriptional regulation of biomass-degrading enzymes in the  
 filamentous fungus Trichoderma reesei  
 JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)  
 MEDLINE 22803314  
 PUBMED 12788920  
 COMMENT Contact: Pamela K. Foreman  
 Genencor Intl  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: Pforeman@genencor.com  
 Seq primer: LT-F1 primer.  
 Location/Qualifiers  
 1..761  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="tric029xo13"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 3 april"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,73e-43 Length: 761  
 Score: 615.50 Matches: 118  
 Percent Similarity: 75.61% Conservative: 37  
 Best Local Similarity: 57.56% Mismatches: 43  
 Query Match: 15.25% Indels: 9  
 DB: 14 Gaps: 4

US-09-917-376-3 (1-740) x CB902332 (1-761)  
 QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 DB TTTTCATCGAAGACAGTCAAGCTCGCGCGCGCGCGCTTCGTCCCGGCATCATCTTC 214  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43  
 DB CATCCCAAGACAAAGGGGTAGCATATGCACGAACAGATATGGGGCTGTACCGCTC 274  
 QY 44 TrpAspAlaAlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGlyTrpAsn 61







```

213 TCTGACCCGCGGAGAGCGAGCTATTGGCCGACGAGTCGGAGACGACACGCTTCACCTTT 272
457 Ala-----AspValThrAlaValProSerThrPheThrPheThrValPheThr 473
273 GCCAGCAGAAACACCTCGGAGCATCCCGAGAGCGTCTGGGCAACGCCACATGGGCC 332
474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySer 493
333 ACCTCGACGAGCGTCGACTACGCGCGGAACTCGGTCAAGAGCGTCTGGCGGTCGGCAAC 392
494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513
393 ACCGCCGCGACCAACAG-----GTGGCCATCTCTGTCGAGCGGCGCG 437
514 AsnTrpPheGlnGlySerGluProGlyValThrGlyValThrThrGlyThrValAlaAspSer 533
438 ACGTGGAGCATCGACTACGGCGGCGACACGTCATGAAACGGCGCACGGTGGCTATTTCG 497
534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
498 GCCAGCGCGCACACGATCCTCTGGTCACCGCTCTGTCGGC-----GTGACGGC 548
554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573
549 TCGCAGTTCAGGGCAGCTTTCCTCGTCTCGAGCCTCGCGCGGCGCGTCAATCGCC 608
574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593
609 TCGGACAAAGAACCAACAGCGTCTTCTACGCGGCTCCGATCGACCTTTTACGTGAGC 668
594 ThrAspGlyValThrPhe 600
669 AAGGACACCGCGACGAGCTTC 689

```

```

RESULT 6
CB907625          782 bp  mRNA  linear  EST 02-JUL-2003
LOCUS             trico02xn19 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION        jecorina cDNA clone trico02xn19, mRNA sequence.
ACCESSION         CB907625
VERSION           CB907625.1 GI:30122283
KEYWORDS           EST.
SOURCE            Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM          Hypocrea jecorina

```

```

REFERENCE
AUTHORS           Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
                  Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
                  Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
                  Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
                  Transcriptional regulation of biomass-degrading enzymes in the
                  filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)

```

```

JOURNAL           22803314
MEDLINE           12788920
PUBMED            Genencor Intl.
COMMENT           Contact: Pamela K. Foreman
                  925 Page Mill Road, Palo Alto, CA 94304, USA
                  Tel: (650) 846-7635
                  Fax: (650) 621-7817
                  Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
1..782
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico02xn19"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"

```

## FEATURES

```

source
CF882065          693 bp  mRNA  linear  EST 31-OCT-2003
LOCUS             trico29xn03 bl1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION        Hypocrea jecorina cDNA clone trico29xn03, mRNA sequence.
ACCESSION         CF882065
VERSION           CF882065.1 GI:38136747
KEYWORDS           EST.
SOURCE            Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM          Hypocrea jecorina

```

/note="vector: PREP3Y, Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

```

ORIGIN
Alignment Scores:      8.7e-26      Length:      782
Pred. NO.:            415.50      Matches:     92
Score:                55.51%      Conservative: 34
Percent Similarity:    40.53%      Mismatches:  90
Best Local Similarity: 10.29%      Indels:      11
Query Match:          14          Gaps:        3
DB:

US-09-917-376-3 (1-740) x CB907625 (1-782)

Qy 377 LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396
Db 108 CGCTTCGGCTCGATGATTGAGTCTCTCGAGATTGACCAACCGCAGCAGCAACCACTGGCTC 167
Qy 397 TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly 416
Db 168 TAGCGCACCGGATGACAAATCTTTGGCGGCCACGATCTCACCACTGGGACACGCGCCAC 227
Qy 417 GlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIle 436
Db 228 AATGTGTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 287
Qy 437 SerProProSerGlyAlaProIleLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456
Db 288 TCTGCACCCCGCGGAAGCGAGCTATTGGCCGAGTCGGAGACGACCAACGGCTTCACCTTT 347
Qy 457 Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr 473
Db 348 GCCAGCAGAAACGACCTCGGACATCGCGCAGACGGTCTGGGCAACGCCACATGGGCC 407
Qy 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySer 493
Db 408 ACCTCGACGAGCGTCGACTACGCGCGGAACTCGGTCAAGAGCGTCTGCGCGTCCGCAAC 467
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513
Db 468 ACCGCCGCGCACCAACAG-----GTGGCCATCTCTGTCGAGCGGCGCG 512
Qy 514 AsnTrpPheGlnGlySerGluProGlyValThrGlyValThrGlyValAlaAlaSer 533
Db 513 ACGTGGAGCATCGACTACGGCGCGCACGTCATGAAACGGCGCACGGTGGCTATTTCG 572
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
Db 573 GCCAGCGCGCACGATCCTCTGGTCCGACCGCTCTGTCGGC-----GTGACGGC 623
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573
Db 624 TCGCAGTTTCCAGGGCGAGCTTTGCTCTCGAGCGTCTCGAGCGTCCCGCGGCGCGTCATCG 683
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593
Db 684 TCGGACAAAGAACCAACAGCGTCTTCTACGCGGCTCCGGATCGGATTCGACCTTTTACGTG 743
Qy 594 ThrAspGlyValThrPhe 600
Db 744 AAGGACACCGCGACGAGCTTC 764

```



Qy 504 HisValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGluProGlyGly 523  
 Db 303 ---GTGCGCATCTCGTCGACGGCGCGGCGACGTCGAGCATCGACTACGGCGCGCACACG 359  
 Qy 524 ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543  
 Db 360 TCCATGAACGCGCGCGACGGTGGCTATTTCGCGCGACGCGACGATCCTCTGTGTCCGCC 419  
 Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563  
 Db 420 GCCTCTCGCGC-----GTGCGCGCTCGGAGTTCAGGCGCATCTTGCCTCCGTC 470  
 Qy 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583  
 Db 471 TCGAGCTGCGCGCGCGCGTTCATCGCTCGGACCAAGAACCAACAGCGCTCTTCTAC 530  
 Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603  
 Db 531 GCCCGCTCGCGATCGACCTTTTACGTGAGCAAGGACACCGCGACGAGCTTC----- 581  
 Qy 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621  
 Db 582 ACGCGCGG---CCCAAGCTGGCGCGCGAGGAGCATCCGGGATATCGCTGCTCACCG 638  
 Qy 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641  
 Db 639 ACCACCGCGCGACGTTGTATGTTCTCGACCGACGTCGGCATATTTCCTCCACAGACTCG 698  
 Qy 642 GlySerSerTrpSerAlaIleThr 649  
 Db 699 GGCACGACCTTTGGCCAAAGTCTCC 722

## RESULT 9

CB898982 803 bp mRNA linear EST 02-JUL-2003  
 LOCUS tricol16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 DEFINITION jecorina cDNA clone tricol16xm14, mRNA sequence.

## ACCESSION

CB898982

## VERSION

CB898982.1 GI:30113640

## KEYWORDS

EST.

## SOURCE

Hypocrea jecorina (anamorph: Trichoderma reesei)

## ORGANISM

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

## REFERENCE

1 (bases 1 to 803)

## AUTHORS

Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,

Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,

Kelley, A.S., Meerman, H.J., Mitchell, T.S., Mitchinson, C.,

Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.

Transcriptional regulation of biomass-degrading enzymes in the

filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

22803314

12788920

Contact: Pamela K. Foreman

Genencor Intl.

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: LT-Fl primer.

Location/Qualifiers

1..803

/organism="Hypocrea jecorina"

/mol\_type="mRNA"

/strain="QM6a"

/db\_xref="taxon:51453"

/clone="tricol16xm14"

/dev\_stage="mycelia"

/clone\_lib="T.reesei"

/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

## FEATURES

## source

RESULT 10  
 LOCUS CF866388  
 DEFINITION tricol06xe08.b1 T.reesei mycelial culture, Version 6 October 2003  
 ACCESSION CF866388  
 VERSION CF866388.1 GI:38121014  
 KEYWORDS EST.  
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,78e-21 Length: 803  
 Score: 361.50 Matches: 88  
 Percent Similarity: 51.61% Conservative: 40  
 Best Local Similarity: 35.48% Mismatches: 103  
 Query Match: 8.96% Indels: 17  
 DB: 14 Gaps: 6

US-09-917-376-3 (1-740) x CB898982 (1-803)

Qy 407 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGly 426  
 Db 67 CACGATTTTCCAACTGGGACACGGCCACATGTTGTTCAATCCAATCACTGTCGACGCGC 126  
 Qy 427 LeuGluGluThrAlaValAsnAspLeuLeuSerProProSerGlyAlaProIleLeuSer 446  
 Db 127 ATCAGGGAATTTCTCGTCCAGGACCTGCGCTTGCACCGCGGGAAGCAGCTATTGGCC 186  
 Qy 447 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 463  
 Db 187 CGAGTCGGAGACGACACACGGTTTACCTTGCACGAGAAACGACCTCGGACATCGCCG 246  
 Qy 464 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483  
 Db 247 CAGACGGTCTGGGCAACGCCACATGGGCCACCTCGACGAGCGTCGACTACGCGCGGAAC 306  
 Qy 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503  
 Db 307 TCGGTCAGAGCGTCTCGCGTTCGGCAACACCGCGCGGCAACAG-----354  
 Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
 Db 355 ---GTGCGCATCTCGTCGACGGCGCGCGAGCATCGATCGATCGCGCGCGACACG 411  
 Qy 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543  
 Db 412 TCCATGAACGCGCGCGACGGTGGCTATTTCGCGCGGCGGCGACGATCCTCTGTGTCCGCC 471  
 Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563  
 Db 472 GCCTCTCGCGC-----GTGCGCGCTCGGAGTTCAGGCGCGAGCTTGTGCTCCGTC 522  
 Qy 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583  
 Db 523 TCGAGCTTCCCGCGCGCGCGTTCATCGCTCGGACCAAGAACCAACAGCGCTTCTTCTAC 582  
 Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603  
 Db 583 GCCCGCTCGGATCGACCTTTTACGTACCAAGGACACCGCGCGAGCTTC-----633  
 Qy 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621  
 Db 634 ACGCGCGG---CCCAAGCTGGCGCGCGGCGAGCATCGGCGATATCGTGTCTCACCCG 690  
 Qy 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641  
 Db 691 ACCACCGCGCGACGTTGTATGTTCTCGACCGGACGTCGGCATATTTCGCTCCACAGACTCG 750  
 Qy 642 GlySerSerTrpSerAlaIleThr 649  
 Db 751 GGCACGACCTTTGGCCAAAGTCTCC 774

CF866388 738 bp mRNA linear EST 31-OCT-2003  
 tricol06xe08.b1 T.reesei mycelial culture, Version 6 October 2003  
 Hypocrea jecorina cDNA clone tricol06xe08, mRNA sequence.  
 CF866388  
 CF866388.1 GI:38121014  
 EST.  
 Hypocrea jecorina (anamorph: Trichoderma reesei)



US-09-917-376-3 (1-740) x CB896354 (1-794)

Qy 422 ProMetVallys---GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer 440  
 Db 58 CCACGCTCCGAGCGCATCGAGGATTCCTCGTCCAGGACCTGGCTTCGACCCGGC 117  
 Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457  
 Db 118 GGAAGCGAGTATTGGCGGAGTCGAGACGCAACACGGCTTCACCTTTGCCAGCAGAAAC 177  
 Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 477  
 Db 178 GACCTCGGACATCGCGACAGCGTTCGGCAGCGCCACATGGGCCCATCGACAGC 237  
 Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSer 497  
 Db 238 GTGACTACCGCGGAACTCGTCAAGAGCGTGTCTCGGTGCGCAACACCGCCGGCAGC 297  
 Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517  
 Db 298 CAACAG-----GTGGCCATCTCGTCCGACGCGCGCGCGACGTGGAGCATC 342  
 Qy 518 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537  
 Db 343 GACTACGCGCGCACACATCATGAACGCGGCGCGTGGCTATTGGCGCGACGCGAC 402  
 Qy 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557  
 Db 403 ACGATCTCTGGTCCGACCGCTCGTCCGCG-----GTGCAGCGCTCGAGTTCAG 453  
 Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577  
 Db 454 GCGAGCTTTCCTCGCTCGAGCGCTCCCGCGCGCGCTCATCGCTCGGACAGAG 513  
 Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597  
 Db 514 ACCAACAGCGCTCTTACTACGCGCGCTCGGATCGACCTTTTACGTCAGCAGACGACCGCGC 573  
 Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616  
 Db 574 AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGCGCAGCGGACGATCCGG 621  
 Qy 617 ---PheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635  
 Db 622 GATATCGCTGCTACCCGACACCGCGCGGACGTTGTATGTCTGACGCGCGCTCGGATA 681  
 Qy 636 TyrHisSerThrAsnGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654  
 Db 682 TTCGCTCCACAGACTCGCGGCGACACCTTTGGCCAAAGTCTCCACCGCGCTGACCAACACC 741  
 Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668  
 Db 742 TACCAGATCGCTCGGTGTGGGTCTCA---GGCTCGAACTGG 780

RESULT 12  
 LOCUS BF072664 546 bp mRNA linear EST 18-OCT-2000  
 DEFINITION NCSM3H73 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5' similar to avicelase III, Aspergillus aculeatus, mRNA sequence.  
 ACCESSION BF072664  
 VERSION BF072664.1 GI:10866169  
 KEYWORDS EST.  
 SOURCE Neurospora crassa  
 ORGANISM Neurospora crassa  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 1 (bases 1 to 546)  
 REFERENCE  
 AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L., Leonard, P.M., Mitchell, J., Amijo, A.M., Bean, L., Blueyes, E., Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R., Ortega, J., Pavlova, I., Pera, J., Todisco, S., Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig, D.O.

TITLE Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa  
 JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)  
 MEDLINE 97435549  
 PUBMED 9290248  
 COMMENT Contact: Natvig, D.O./Nelson, M.A.  
 Department of Biology  
 University of New Mexico  
 Castetter Hall, Albuquerque, NM 87131, USA  
 Tel: 505 277 3411  
 Fax: 505 277 0304  
 Email: ngp@biology.unm.edu.  
 Location/Qualifiers  
 1. 546

## FEATURES

source

/organism="Neurospora crassa"  
 /mol\_type="mRNA"  
 /strain="74-OR23-IV A (FGSC 2489)"  
 /db\_xref="taxon:5141"  
 /clone="SM3H7"  
 /sex="Mating type A"  
 /tissue\_type="Mycelium"  
 /dev\_stage="Mycelium"  
 /lab\_host="E. coli"  
 /clone\_lib="Subtracted Mycelial"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; 2% sucrose for 24 hours. cDNA directionally cloned into pBluescript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA). Previously identified highly expressed clones were subtracted from this library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3 29e-17 Length: 546  
 Score: 314.50 Matches: 72  
 Percent Similarity: 54.26% Conservative: 30  
 Best Local Similarity: 38.30% Mismatches: 75  
 Query Match: 7.79% Indels: 11  
 DB: 10 Gaps: 3  
 US-09-917-376-3 (1-740) x BF072664 (1-546)

Qy 358 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLys 377  
 Db 5 TCGGCACGAGCTGGATCGAGACTGGCTCTTTTCAG-----GATACCAAGCAC 55  
 Qy 378 LeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeuTyr 397  
 Db 56 CTCGGCTGGATGATCGAATCCCTCGAGATCAACCCCTCGACAGCGACCATTCGCTCTAC 115  
 Qy 398 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGln 417  
 Db 116 GCCACCGGTCTCAGTGTTCAGCGCGCCACGACCTGACCAAGTGGGACACCGTCCACAC 175  
 Qy 418 IleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSer 437  
 Db 176 GTGACCATCCAGTCTCTGGCCGCTTGGCATCGAAGAAATGGCTCTCGTCTCGCTCC 235  
 Qy 438 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThr----- 455  
 Db 236 GCCCCGGTGGTCCGAGCTTCTCGCGCGCTCGGTGACGACTGCGCGCTTCACCTTCAAG 295  
 Qy 456 ---HisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 474  
 Db 296 TCCAGCTCCGACCTCGGCACCTCTCTCCAAAGACCTGGATGACGCCCGCCAGTGGCCGAGC 355  
 Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 494  
 Db 356 TCCGGCGAGCTCGACTAGCCGCGCAACAGCCCGCAAGCTGGTGGCATCGCTCGGCTCGGC 415  
 Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514  
 Db 416 TCCGGCGCGCAGCAG-----GTGGCGGTATCTCTCGGACGCGCGCGCTCG 460

```
QY 515 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 534
Db 461 TGGCATGCTCAACAGCGCAGCACACACCAAGAGCAGCGGCGCTACTCTGCC 520
QY 535 AspGlySerArgPheValTIPAla 542
Db 521 GACGCCGACCATCGTCTGTCT 544

RESULT 13
BU639045 389 bp mRNA linear EST 06-MAY-2003
LOCUS mcgw011xd18.f RCW Lambda Zap Express Library Magnaporthe grisea
DEFINITION cDNA clone mcgw011xd18 5', mRNA sequence.
ACCESSION BU639045
VERSION BU639045.2 GI:30397224
KEYWORDS EST.
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 389)
AUTHORS Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G.,
Bhatteirai,K. and Dean,R.A.
TITLE Expressed sequence tags from the rice blast fungus, Magnaporthe
grisea
JOURNAL Unpublished (2002)
COMMENT On Sep 30, 2002 this sequence version replaced gi:23351371.
Contact: Ebbole DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
Tel: 979 845 4831
Fax: 979 845 6483
Email: d-ebbole@tamu.edu
Chromatogram file of this sequence is available, see contact
person; Best nr hit (April. 22, 2003) gb|EAA29333.1| hypothetical
protein [Neurospora crassa] 206 4e-53
PCR Primers
FORWARD: T3 primer
BACKWARD: T7 primer
Plate: mcgw011 row: D column: 18
Seq primer: T3.

FEATURES             source
    source
    1..389            Location/Qualifiers
                        /organism="Magnaporthe grisea"
                        /mol_type="mRNA"
                        /strain="CP987"
                        /db_xref="taxon:148305"
                        /clone="mgcw011xd18"
                        /sex="Mat1-1 hermaphrodite"
                        /tissue type="Mycelium"
                        /dev stage="Day 5 post-inoculation"
                        /clone_lib="RCW Lambda Zap Express Library"
                        /notes="Vector: pBluescript excised from Lambda Zap
                        Express; Site 1: EcoRI; Site 2: XhoI; Day 5
                        post-inoculation mRNAs prepared from Magnaporthe grisea
                        grown at 23C in the dark with constant gyratory shaking
                        100 rpm in Vogel's minimal medium containing 0.5% isolated
                        rice cell walls as the sole carbon source. Library
                        provided by Sheng-Cheng Wu. Sequences were processed by
                        one of two methods. Where a full-length alignment to the
                        M. grisea genome sequence was available, the EST sequence
                        was trimmed according to the alignment, otherwise sequence
                        quality was assessed using phredPhrap version 991019 and
                        trimmed according to phd files (0.05) and for vector
                        segs."
```

ORIGIN

Alignment Scores: 9.49e-17 Length: 389  
Pred. No.: 306.50 Matches: 55  
Score: 66.67% Conservative: 27  
Best Local Similarity: 44.72% Mismatches: 38

```
Query Match: 7.59% Indels: 3
DB: 13 Gaps: 1
US-09-917-376-3 (1-740) x BU639045 (1-389)

QY 334 GlyAlaThrTIPThrArgIleTIPAspTIPThrSerTyPProAsnArgSerLeuArgTyr 353
Db 5 GGGGCTACATGTCGAGACTTTGGGAGTCGACGATATCCGACATGAACCACTACTAC 64
QY 354 ValLeuAspIleSerAlaGluProTIPLeuThrPheGlyValGlnProAsnProVal 373
Db 65 TCGATCAACACACAGACAGCGCCCTGGATAGAGTCTGGC-----CATCTTCGAGG 115
QY 374 ProSerProLysLeuGlyTIPMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393
Db 116 GATTTCGAGCGCTTGGTTGGATGATCGAGGCCCTTGAAATCGACCCCTGTCGATCCTGAC 175
QY 394 ArgMetLeuTyGlyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTIPAsp 413
Db 176 CATGGCTTTACGGAACCTGGCCTCTCGATCTTTGGGGGACATGACTTTGACCAATGGGAC 235
QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433
Db 236 ACGGTTCAACATATCGATTCACTTTGGCCGATGGTGTGGAAGAACTGCGAGTTCTG 295
QY 434 AspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453
Db 296 GAGATGGCTTCAGCGCCGCGAGGCTCCGAACACTACTGGCCGAATTCGAGACATTACTGGA 355
QY 454 PheThrHis 456
Db 356 TTCACATAC 364

RESULT 14
CF875815 713 bp mRNA linear EST 31-OCT-2003
LOCUS trico39xk06.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico39xk06, mRNA sequence.
ACCESSION CF875815
VERSION CF875815.1 GI:38130497
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 713)
AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
Dean,R.A.
TITLE Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: Tr-F1 primer.
Location/Qualifiers
    1..713            /organism="Hypocrea jecorina"
                        /mol_type="mRNA"
                        /strain="QM6a"
                        /db_xref="taxon:51453"
                        /clone="trico39xk06"
                        /dev stage="mycelia"
                        /clone_lib="T.reesei mycelial culture, Version 6 October
                        2003"
                        /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                        culture grown from 24 hrs to 6 days with varying Carbon
                        and Nitrogen sources and concentrations."
```

```

ORIGIN
Alignment Scores:
Pred. No.: 1.42e-14 Length: 713
Score: 286.50 Matches: 81
Percent Similarity: 48.79% Conservative: 40
Best Local Similarity: 32.66% Mismatches: 106
Query Match: 7.10% Indels: 21
DB: 14 Gaps: 10

US-09-917-376-3 (1-740) x CF875815 (1-713)
Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 10 GACCTCGGTATCATCGCGAGTGTGTGGCAACGCCACATCGGCCACCTCGACGAGC 69
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleValAlaGlySerPheAspProSer 497
Db 70 GTCGACTACGCGGGAACCTCGTCAAGAGCGTGTGGCGTGGCAACACCGCGGCGACG 129
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 130 CAACAG-----GTGGCCATCTCTCGCGAGCGCGCGCGGCGGAGCATC 174
Qy 518 GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537
Db 175 GACTACGCGCGGACACATCATCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 234
Qy 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValThrValAlaValGlyPheGly 557
Db 235 ACGATCTCTGTGTCAGCGCGCTCTGTCGCG-----GTGAGCGCTCGAGTTCCAG 285
Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
Db 286 GCGAGCTTGTCTCGTCTGAGCTGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 345
Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
Db 346 ACCAACAGCGTCTTACGCGCGGCTCCGATCGACCTTTTACGTCAGCAAGACACCGCGC 405
Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616
Db 406 AGCAGCTTC-----ACGCGCGGG---CCCAAGCTGGCGGCGGCGGCGGCGGCGGCGG 453
Qy 617 ---PheHisAlaValProGlyLysGlyValAspLeuTrpLeuAlaAlaSerSerGlyLeu 635
Db 454 GATATCGCTGTCTACCGGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513
Qy 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654
Db 514 TTCGCTCCACAGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 674
Db 574 TACCAGATCGGCTGGGTGTGGGTCA---GGCTCGAACTGG---AACCTGATGCTTC 627
Qy 675 GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr 692
Db 628 GGCAC-----GGCCGCTAGGGGCTCGCTCTACGCCAGTGAGAGACGCGGCGCTCC 681
Qy 693 TrpValLeuIleAsnAspAspGln 700
Db 682 TGGACGACATTTAGGGCTCCAG 705

RESULT 15
CB904767
LOCUS
DEFINITION
trio039xk06 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trio039xk06, mRNA sequence.
ACCESSION
CB904767.1 GI:30119425
VERSION
CB904767.1
KEYWORDS
Hypocrea jecorina (anamorph: Trichoderma reesei)

```

```

ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 782)
AUTHORS
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL
MEDLINE
PUBMED
22803314
12788920
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
FEATURES
Location/Qualifiers
1..782
source
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trio039xk06"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.: 1.65e-14 Length: 782
Score: 286.50 Matches: 81
Percent Similarity: 48.79% Conservative: 40
Best Local Similarity: 32.66% Mismatches: 106
Query Match: 7.10% Indels: 21
DB: 14 Gaps: 10

US-09-917-376-3 (1-740) x CB904767 (1-782)
Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 79 GACCTCGGTATCATCGCGAGTGTGTGGCAACGCCACATCGGCCACCTCGACGAGC 138
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleValAlaGlySerPheAspProSer 497
Db 139 GTCGACTACGCGGGAACCTCGTCAAGAGCGTGTGGCGTGGCAACACCGCGGCGACG 198
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 199 CAACAG-----GTGGCCATCTCTCGCGAGCGGCGGCGGCGGCGGCGGCGGCGG 243
Qy 518 GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537
Db 244 GACTACGCGCGGCGACACATCGTCCATGAACGCGGCGGCGGCGGCGGCGGCGGCGG 303
Qy 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
Db 304 ACGATCTCTGTGTCAGCGGCTCTGTCGCG-----GTGCGGGCTCGGAGTTCCAG 354
Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
Db 355 GGCAGCTTGTGCTCGTCTCGAGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414
Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
Db 415 ACCAACAGCGTCTTCTACGCGGCTCTCGGATCGACCTTTTACGTACGACGAGACACCGG 474
Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616

```



475	Db	AGCAGCTTC-----ACCGCGGG---CCCAAGCTGGCGACGCGGACGATCCGG	522
617	Qy	--PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerglyLeu	635
523	Db	GATATCGCTGCTCACCCGACACCGCGGGCAGCTTGATGCTCGACCGACGTCGCGCAT	582
636	Qy	TyrHisSerThrAsnGlyGlySerTrpSerAlaIle---ThrGlyValserSerAla	654
583	Db	TTCCGCTCCACAGACTCGGGCAGCAGCTTGGCCCAAGTCTCCACGCGCTGACCAACACC	642
655	Qy	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	674
643	Db	TACCAGATCGCCCTGGGTGTCGGGTCA---GGCTCGAACTGG---AACTGTATGCCTTC	696
675	Qy	GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspCysgelyThrThr	692
697	Db	GGCACC-----GGCCCTCAGGGGCTCGCTCTACGCCAGTGGAGACAGCGGCGCTCC	750
693	Qy	TrpValLeuIleAsnAspGln	700
751	Db	TGGACGGACATTCAGGGCTCCACG	774

Search completed: May 12, 2004, 11:38:55  
Job time : 6964.51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 184.798 Seconds

(without alignments)

2222.227 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGFVD.....YITNGRGIVYDIGGAPSG 740

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh

-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1429

-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPECL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09917376 -CGN1\_1116 -runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DE TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:

1: /cgn2\_6/prodata/2/ina/5A COMB.seq:

2: /cgn2\_6/prodata/2/ina/5B COMB.seq:

3: /cgn2\_6/prodata/2/ina/6A COMB.seq:

4: /cgn2\_6/prodata/2/ina/6B COMB.seq:

5: /cgn2\_6/prodata/2/ina/6C COMB.seq:

6: /cgn2\_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	227.5	5.6	3129	US-09-252-991A-13873	Sequence 13873, A
C 2	227.5	5.6	8211	US-09-252-991A-13656	Sequence 13656, A
C 3	203	5.0	4188	US-09-252-991A-13774	Sequence 13774, A
C 4	201.5	5.0	4131	US-09-252-991A-13773	Sequence 13773, A
C 5	192	4.8	4411529	US-09-103-840A-1	Sequence 1, Appli
C 6	184	4.6	4403765	US-09-103-840A-2	Sequence 2, Appli
C 7	181.5	4.5	11679	US-09-328-352-1377	Sequence 39, Appl
C 8	170.5	4.2	25165	US-09-453-702B-39	Sequence 2, Appli
C 9	170	4.2	4403765	US-09-103-840A-2	Sequence 1, Appli
C 10	169	4.2	4411529	US-09-103-840A-1	Sequence 1, Appli
C 11	164	4.1	2319	US-09-252-991A-13875	Sequence 13875, A
C 12	163	4.0	5741	US-07-706-699-4	Sequence 4, Appli

13	163	4.0	5741	1	US-07-998-931-4	Sequence 4, Appli
14	160.5	4.0	47981	4	US-09-679-279-1	Sequence 1, Appli
15	159.5	4.0	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
16	159.5	4.0	10419	4	US-09-408-020-3	Sequence 3, Appli
17	159.5	4.0	42432	4	US-09-408-020-2	Sequence 2, Appli
18	157.5	3.9	2067	4	US-09-489-039A-2642	Sequence 2642, Ap
19	156.5	3.9	1860	4	US-09-252-991A-9781	Sequence 9781, Ap
20	155	3.8	2235	1	US-08-418-782-1	Sequence 1, Appli
21	155	3.8	2235	2	US-08-228-662-1	Sequence 1, Appli
22	155	3.8	2235	2	US-08-852-219-1	Sequence 1, Appli
23	154	3.8	3900	4	US-09-023-655-1420	Sequence 1420, Ap
24	153	3.8	2625	6	5457037-4	Patent No. 5457037
25	153	3.8	3336	6	5457037-1	Patent No. 5457037
26	153	3.8	18034	4	US-09-266-965-75	Sequence 75, Appl
27	152.5	3.8	6416	3	US-09-136-574A-2	Sequence 2, Appli
28	151	3.7	1989	4	US-09-252-991A-15371	Sequence 15371, A
29	151	3.7	2637	4	US-09-252-991A-15371	Sequence 15403, A
30	150.5	3.7	2058	4	US-09-252-991A-15403	Sequence 15462, A
31	149.5	3.7	2253	6	5457037-2	Patent No. 5457037
32	149	3.7	2852	4	US-09-056-556-203	Sequence 203, App
33	149	3.7	2852	4	US-09-072-596-198	Sequence 198, App
34	149	3.7	2852	4	US-09-072-967-203	Sequence 203, App
35	149	3.7	3120	4	US-09-169-768-29	Sequence 29, Appl
36	147.5	3.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
37	147.5	3.7	44377	2	US-08-804-198-1	Sequence 1, Appli
38	146	3.6	87563	4	US-09-453-702B-57	Sequence 57, Appl
39	145	3.6	2163	3	US-09-296-284-22	Sequence 22, Appl
40	145	3.6	2185	1	US-08-173-508-3	Sequence 3, Appli
41	145	3.6	2185	2	US-08-285-310-3	Sequence 3, Appli
42	145	3.6	2185	3	US-08-951-742-3	Sequence 3, Appli
43	145	3.6	2265	3	US-09-296-284-1	Sequence 1, Appli
44	145	3.6	4830	3	US-09-296-284-7	Sequence 7, Appli
45	144.5	3.6	1164	4	US-09-252-991A-5391	Sequence 5391, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-13873/c

; Sequence 13873, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13873

; LENGTH: 3129

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13873

Alignment Scores:	7.76e-11	Length:	3129
Pred. No.:	227.50	Matches:	218
Score:	31.30%	Conservative:	90
Percent Similarity:	22.15%	Mismatches:	360
Best Local Similarity:	5.64%	Indels:	319
Query Match:	4	Gaps:	52
DB:			

US-09-917-376-3 (1-740) x US-09-252-991A-13873 (1-3129)

Qy 20 AspGlyIleValPheAenGluGlyAla---ProGlyIleLeuTyValArgThrAspIle 38

Db 2993 AACGGCGTCGTCATCAGCGGACCGCGGAGCGCGTCCACCGTACCGTCACCGATGCC 2934

Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58  
Db 2933 GGC-----GGCAACCCGATGAGCGCAGGTCAACCGCGAC 2901  
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
Db 2900 GGCAGCGGCACTGGAGCTTCACCGCGGGCACCGCGCGGCAACGGCAAGGTGATGTC 2841  
Qy 70 SerIleAlaAlaAspProIle-----AsnThrAsnLysValTrpAla-----Ala 84  
Db 2840 GCCAGCGCCACCGACCGCGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2781  
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu-----100  
Db 2780 GTGGCGCGCGCGCGCGCGGTGATCGATCCGAGCAACCGCGCACCATCATCGCGCACCGCG 2721  
Qy 101-----ArgSer 102  
Db 2720 GAGCGCGCGCGCGCGCGGTGATCCTCACCGACGGCAACCGCGCAACCGGATCGCGGCAACCC 2661  
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro-----LeuProPheLysLeuGly-----119  
Db 2660 GCCGACGGCAGCGCACTGGACCTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2601  
Qy 120-----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
Db 2600 GTCAACCG 2541  
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
Db 2540 GAGCGGTGGCGCGCGCAACCGCTGTGTGTCATCCGAGCAACCGCGCAACCTGTCTCAACGGT 2481  
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
Db 2480 ACCG 2421  
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
Db 2420 ACCACCG 2364  
Qy 168 ProAspValGlyThrTyrlle-----AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186  
Db 2363-----GGCACCGGTGGTCAACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2316  
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln-----204  
Db 2315-----GCTCCCGCTACCACGCGGTGGATTCCTGCTGCGCTGATCCCGCGAGGTGAT 2262  
Qy 205---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe-----221  
Db 2261 CCGAGCAACGGTTCGGTGTATCAGCGGCACCGCGCGCGCGCGCGCGCGCGCGCGCG 2202  
Qy 222-----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 2201 GATGGCAACGGCAACCGGATTTGCCAGGTCAACCGCGCGCGCGCGCGCGCGCGCGCG 2142  
Qy 233 ValProGlyAlaPro-----237  
Db 2141 ACTCCAGGCATCCGCTGCGGATGGACCGGTGTCACCGGTGGCGCGCGCGCGCGCGCG 2082  
Qy 238-----ThrGlyPheIleProHisLysGlyVal 246  
Db 2081 AATGTGCACAGTCGCGCGCGGTGATCACTGTGTGATGGCGTGGCGCGCGCGCGCGCG 2022  
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr---265  
Db 2021 ATCGATCCGAGCAACCGCACCGAGATAAGCGGTACCGCGGAGCGCGCGCGCGCGCGCG 1962  
Qy 266-----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282  
Db 1961 CTCACG 1902  
Qy 283 ThrArgIleSerProValPro-----SerThrAsp 292

Db 1901 ACGTTTACCCCGCGCACCCCGCTGGCCCAACCGCACCGGTGATCAACCGCGCGCGCGCGAC 1842  
Qy 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro-----310  
Db 1841 CCGCGCGCAATACACGCGGTCCGCGCGCGGTCAACCGGTGATCGCATCGCGCGCGCGCG 1782  
Qy 311-----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 1781 CCGGTGATCAATCCGAGCAACGAGTCTCATCAGCGGTACCGCGGAACCGCGCGCGCG 1722  
Qy 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 1721 GTGATCTCTC-----ACCGAGCGCAACCGCAACCGCATCGCGCGCGCGCGCGAC 1671  
Qy 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerIle 351  
Db 1670 GCGAGCGCGAAGTGGGCTTTCACGCGCGCGCGCGCGGTGGCCAAATGGGACGGTGTATCAAT 1611  
Qy 351 uArgTyrValIleuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371  
Db 1610 GCGCTGG-----CCGAGGACCGCGCGCGCGCAACACAGCAGTC-----CC 1572  
Qy 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs 391  
Db 1571 ACCAGCGCCA---CCGTGCGACTCGTGGCGCGAGCA-GCCCGCGGTGATCGATCCGAGCAA 1516  
Qy 391 nSerAspArgMetLeuTyrGlyThr-----GlyValThrLeuTyrAlaThrAsnAs 408  
Db 1515 CGGTAGC---GTGATCGCGGTACCGCGAGCGTGTGTCACGGTGTCTCTCACC-----1464  
Qy 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG 428  
Db 1463-----CACGGCAACGGCAAC-----CCGATCGCGCGCGGTCAACCGC 1429  
Qy 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu-----444  
Db 1428 CGATGGCAGCGCAACTGGAGCTTCACGCC-----GGCACCGCGCTGTCCAATGGCAC 1375  
Qy 445---IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 1374 GTGGTCAATCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1315  
Qy 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal--478  
Db 1314 GTGGACTCGGTGGCG 1255  
Qy 479 -AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 498  
Db 1254 CCGTACCGCGGAGCGCGGTGCGCGGTGATCTCTACCGATGGCGCGCGCAACCGCGATCG 1195  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----515  
Db 1194 CCAGGCC-----ACCGCGGATGGCAGCGCGCAACTGGAGCTTAC 1156  
Qy 516-----PheGlnGlySerGluProGlyG 523  
Db 1155 CCGCGGCACCGCGTGGCCCAACGGCACCGGTGATCAATCGGTGGCGCGCGCGCGCGCG 1096  
Qy 523 yValThrThrGlyGlyThr-----529  
Db 1095 CAATACCG 1036  
Qy 530-----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542  
Db 1035 CAACCCGAGCAACCGCAGCGGTGATCGCGGTGTACCGCGGAACCGCGCGCGCGCGCG 976  
Qy 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 975 CACCG 919  
Qy 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer-----574

Db 918 GAGCTTACGCGCCGACGCGCGTGGCCCAACGGCTCGGTGATCAATGCGCTGGCCGACGA 859  
Qy 575 -----AspArgValAsnProLysTh 581  
Db 858 CGCCCGCGGCAACACCGAGCGCGCGCCGAGCAGCAGCAGCGGTGGACTCGGTAGCCCGCGCAC 799  
Qy 581 r-----PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
Db 798 CCGCGTGTCTCGATCCGAGCAACCGGTACGCTGATCAGCGGTACCGCGGAAGCCGCGGCCAC 739  
Qy 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS 610  
Db 738 GGTGATCTCTACCGAGCGCGGCG-----GGCAACCCGATACGCGAGCCAGCCGCGATGG 685  
Qy 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeuA 630  
Db 684 CAGCGCAACTGGAGCTTCACTCCGGGCACA-CCGCTGACCAACGCGACGCGTATCAATG 626  
Qy 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
Db 625 CGGTGGCCGAGGACCGCGCGGCAACACCGCGGTCCGCTCAGCACCACAGTGGAGCGGG 566  
Qy 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT 668  
Db 565 TGGCCCCCGCCACCCCGGTGATCGAGCGAGCAATGGTGTCAAACTCAGCGGCACCGCG 506  
Qy 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679  
Db 505 AACCCGGGTCCGGGTGATCTCACCAGTGGCAATGGCAACCGGATCGGCCAGACCTCG 446  
Qy 680 -----ValT 681  
Db 445 CCGAGGTAGCGGCAACTGGACCTTCACACCGGGCAGCGCGTGGCCACGCGCGTGG 386  
Qy 681 hrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnH 701  
Db 385 TCAACGCGTGGCCCGGAGGACCGCGCGGCAATACCAG-----CGTCCGCGCAGC 336  
Qy 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
Db 335 ACCACGCGTGGATACGCTGGCGCGCGCCAGCGCGGTGATCAATGCCAGCAACGCGCAG 276  
Qy 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738  
Db 275 ATCAGCGGACCGCGAGGTGGCGCGCGCAAGTGTCTCACCAGCAGCGCAACCGCG 216  
Qy 739 SerGly 740  
Db 215 ATCGCG 210

RESULT 2  
US-09-252-991A-13656  
; Sequence 13656, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13656  
; LENGTH: 8211  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13656  
Alignment Scores:

Pred. No.: 3,45e-10 Length: 8211  
Score: 227.50 Matches: 218  
Percent Similarity: 31.30% Conservative: 90  
Best Local Similarity: 22.15% Mismatches: 360  
Query Match: 5.64% Indels: 319  
DB: 52  
US-09-917-376-3 (1-740) x US-09-252-991A-13656 (1-8211)  
Qy 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyrValA-gThrAspIle 38  
Db 3574 AACGGGTGTATCATAGCGGACCGCGGCGGTGGCCACCGGTGACCCCTCACCGATGCC 3633  
Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58  
Db 3634 GGC-----GGCAACCCGATAGGCGAGGTACCGCCGAC 3666  
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
Db 3667 GGCAGCGGCAACTGGAGCTTCACGCGGGCACGCCCGGCCAACCGCACCGGTATCGTC 3726  
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84  
Db 3727 GGCAGCGGACCGCACCGCGGCAATACCGCGCGGCGGCGCCACCGTGGACGCG 3785  
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100  
Db 3787 GTGGCGCGCGCGCGCGGTGATCGATCCGAGCAACGGCACGACCATCAGCGGCACCGCG 3846  
Qy 101 -----ArgSer 102  
Db 3847 GAGCGCGGCGCAAGGTGATCTTCACCGACGCAACCGCAACCGATCGGCGAAACCAACC 3906  
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119  
Db 3907 GCGGACGCGGACGCGCAACTGGACCTTCACCGCGCGCACCGCGTGGCCACGCGCAGGTG 3966  
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
Db 3967 GTCAACGCGGTGGCCCGGACCGCTTGGGGCAATACCGCGCGCGGAGGAGCAGCTACCGTG 4026  
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
Db 4027 GACGCGGTGGCGCGGAACACGCGCTGTGTCAATCCGAGCAACCGCAACCTCTCAACGGT 4086  
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
Db 4087 ACCGCGGAGCGCGGCGAGCACCGTGCTTGACCGACGCAACCGCAACCGATCGGCAG 4146  
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
Db 4147 ACCACCGCGATGGCGAGCGGCAACTGGAGCTTCAGCGCGCGTCCGCACTACCCACAC--- 4203  
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrGlyTyrGlnSerAsp 186  
Db 4204 -----GGCACCGTGGTCAACGTGACCGCGAGCGACGCGCGCGCAATACCGAC--- 4251  
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln----- 204  
Db 4252 -----GCTCCCGCTACCGAGCGGTGGATTCCTCGCTGCGTGCATCCCGCAGGTGGAT 4305  
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe----- 221  
Db 4306 CCGAGCAACCGGTTCGGTGTATCAGCGGCGACCGCGGACCGCGGCAACCATCATCATCACC 4365  
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 4366 GATGGCAACCGCAACCGATTGGCCAGGTTCACCGCGCGACGCGCAGCGGTAACTGGTCTTC 4425  
Qy 233 ValProGlyAlaPro----- 237  
Db 4426 ACTCCAGGCATCCGCTGCGCGGATGGCACCGGTGTCAACGTGTCAACGTGTGGCGCGCAGCCCAAGC 4485

QY 238 -----ThrGlyPheIleProHisLysGlyVal 246  
Db 4486 AATGTGACAGTCGCGCGGGTGTATCACTGTGTGATGGCGCGCGCGCGGTG 4545  
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr--- 265  
Db 4546 ATCGATCCGAGCAACCGCACCGAGATTAAGCGGTACCGCGGAGGCGCGCGGTGATC 4605  
QY 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282  
Db 4606 CTCACCGATGGCGGCGCAACCGCATCGGCAGGCGCACCGCCAGCGCGCAACTGG 4665  
QY 283 ThrArgIleSerProValPro-----SerThrAsp 292  
Db 4666 ACCTTCACCCCGGCGCACCGCTGGCGCAACCGCATCGATCAACGCGGTGGCGCCAGGAC 4725  
QY 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310  
Db 4726 CCGGCGGCAATACCAAGCGGTCCGGCCAGCGTCAACCGTCAATCCATCGCCCGCGCGG 4785  
QY 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 4786 CCGGTGATCAATCCGAGCAACCGAGTTCATCAGCGGTACGCGGAAGCGCGGCGCACG 4845  
QY 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 4846 GTGATCCTC-----ACCAGCGCAACCGCAACCGCATCGCGCAGGTCAACCGCGCAC 4896  
QY 339 ArgIleTrpAspTrpThr---SerTyrProAsnArg-----SerLeu 351  
Db 4897 GCGAGCGCAAGTGGGCTTTCACGCCCGCCACCGCTTGGCCAATGGCAACCGGTGATCAAT 4956  
QY 351 uArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnPr 371  
Db 4957 GGCTGG-----CCAGAGCGCGCGCGCAACACGAGTCT-----CC 4995  
QY 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs 391  
Db 4996 ACCAGCGCCA---CGGTGACTCGTGGCGCCAGCA-GCCCGGTGATCGATCGAGCA 5051  
QY 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408  
Db 5052 CGGTAGC---GTGATCGCGGTACCGCGAGGTGGTGGCCAGCGTATCCTCACC----- 5103  
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuCl 428  
Db 5104 -----GACGGCAACGGCAAC-----CCGATCGCGCAGGTCAACCGC 5138  
QY 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 444  
Db 5139 CGATGGCAGCGCAACTGGAGCTTCACGCC-----GGCAGCGCGCTGTCCAATGGCAC 5192  
QY 445 -----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 5193 GGTGTCAATGCGTGGCGCCAGGAGCGTCCCGGCAACACCGCGCGCGCGCACCAAC 5252  
QY 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478  
Db 5253 GGTGACTCGGTGGCGCGCGCGCGCGGTGATCGACCGCAACCGCGCGGTGATCGC 5312  
QY 479 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 5313 CGGTACCGCGGAAGCGGTGGCGAGCGGTGATCTCACCAGTGGCGGCGCAACCGCATCGG 5372  
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 515  
Db 5373 CCAGGCC-----ACCGCGATGCGAGCGCAACTGGAGCTTCAC 5411  
QY 516 -----PheGlnGlySerGluProGlyG1 523  
Db 5412 CCGGCGCACCGCGTGGCGCAACCGCACCGGTGATCAATCGGTGGCGCCAGGATCGCGCGG 5471  
QY 523 yValThrThrGlyGlyThr----- 529

Db 5472 CAATACACGCGCGCGCACCGGTGGACCGGTGGCCCGCGCGCACCGCGCGTGGT 5531  
QY 530 -----ValAlaIleSerAlaAspGlySerArgPheValTrpAl 542  
Db 5532 CAACCCGAGCAACCGCACCGGTGATCGCGGTATCCGCGGAAGCGCGCGTGTCT 5591  
QY 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 5592 CACGAGCGGCGCGCAACCGCATCGGCGCAGGTCAACCGCGCACCGCGCGCAAC--TG 5648  
QY 560 pAlaIleSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574  
Db 5649 GAGTTTCACCGCGCGCACCGCGTGGCGCAACCGGTTCGTGATCAATGCTGGCGCCACGA 5708  
QY 575 -----AspArgValAsnProLysTh 581  
Db 5709 CGCGCGCGCAACACACGAGCGCGCGCGCACCGGTGACTCGGTAGCCCGCGCAC 5768  
QY 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
Db 5769 CCGGTGCTCGATCCGAGCAACCGGTACGCTGATCAGCGGTACCGCGAAGCGCGCGCAC 5828  
QY 594 -----ThrAspGlyGlyValThrPheGlnProValAlaIleAlaGlyLeu-ProSers 610  
Db 5829 GGTGATCCTCACCGACCGCGC-----GCCAACCCGATACCGGCGCACCGCGCGTGG 5882  
QY 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 5883 CAGCGCAACTGGAGTTTCACTCCGGGCACA-CGGCTGACCAACCGCACCGGTGATCAATG 5941  
QY 630 lAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
Db 5942 CGGTGCGCGAGCGCGCGCGCAACACCGCGGTCCGGTCAGCACCACTGAGACGCGG 6001  
QY 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT 668  
Db 6002 TGCGCGCGCGCACCGCGGTGATCGACCGGCAATGGTGTCTCAACTCAGCGCGCACCGCG 6061  
QY 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679  
Db 6062 AACCGCGGTCCGGGTGATCCTCACCGATGGCAATGGCAACCGCATCGCGCACCGCTCG 6121  
QY 680 -----ValT 681  
Db 6122 CCGACGTAGCGCAACTGGACCTTCACCGCGCACCGCGCTGGCCAAACCGCACCGGTG 6181  
QY 681 hrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnH 701  
Db 6182 TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCAATACCAAG-----CGGTCCGCGCGC 6231  
QY 701 ieGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
Db 6232 ACCCGGTGATACGTGGCGCGCGCGCGCGCGGTGATCAATCCACCAACCGCACCGGTG 6291  
QY 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyValPro 738  
Db 6292 ATCAGCGCGCACCGCGAGGTGGCGCGCAAGTGTCTCACCAGCGCAACCGCACCGCG 6351  
QY 739 SerGly 740  
Db 6352 ATCGCG 6357

## RESULT 3

US-09-252-991A-13774/c  
; Sequence 13774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

Qy	261	ThrGlyGlyProTyr-----	AspGlySerSerGlyAspValTrpLysPheSerVal	277
Db	1821	CGGGCAGCAGCGTGACCTTGACCGATGGCAACCGCAACCCGATCGGCAGGTCAACGCC	1762	
Qy	278	Thr---SerGlyThrTrpThrArgIleSerProValPro-----	289	
Db	1761	GACGGCAGCGGCAACTGGAGCTTCACCCCGTCACCGCGGTGGCGGATGGAAACCGTGGTC	1702	
Qy	290	-----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp	306	
Db	1701	AACGCCACGGCCACCGATCCGGCGGGCAACACACGCGGCAGGGCAGCACCGTCGAT	1642	
Qy	307	ArgGlnHisProAsnThrIleMetVal-----	322	
Db	1641	GGCGTGGCGCGCAGCACCGCAGCCGTCACCTGAGCAACGGCAGCAGCCTCAGCGGCAC	1582	
Qy	323	ProAsp-----ThrIleIlePheArg-----	329	
Db	1581	CGCGAACCGCGCAGCAGCGGTGATCTCTCACGACGGCAACGGCAATCCGATCGCGAGGTC	1522	
Qy	330	SerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArg	349	
Db	1521	ACGCCGACGGCAGCGGCAACTGGACC-----TACACCCCGT--CCACGCCGA	1476	
Qy	350	SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro	369	
Db	1475	TCG-----CCAACGGCACCGTGGTCAACGTGGTGGCCCGCAGGACG	1437	
Qy	370	AsnProProValPro-----	386	
Db	1436	CCGCCGCAATAGCAGCGCGGGCGCACCGCTCACCGTGGAC-TCGAAGACCCCGGGCGCT	1378	
Qy	387	-----IleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----	401	
Db	1377	CCGGTGGTCAACCCGAGCAACGCGC--ACCACCGCTCAGCGGCACCGCGAGCGGGCGCT	1321	
Qy	402	ThrLeuTyrAlaThrAsn-----	414	
Db	1320	ACCGTGAGCTGACGACGGCAACGGCAACCCGATGGCCAGGTCAACGCC--GACGGC	1264	
Qy	415	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp	434	
Db	1263	AGTGGCAACTGGAGCTTCACACCGGCAGCGCGCTGGCCAAACGGCACCGTGGTCAACGCC	1204	
Qy	435	LeuIleSerProSerGly-----	451	
Db	1203	ACGCCCAGCGACCGCAGCGGCAATACAGCGCTCCCGGCCAGCACCGGTGACTCGGTG	1144	
Qy	452	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	471	
Db	1143	-----GGCGCGCGCGCGCGTGGTCAATCCGAGCAACCGAGGTG	1105	
Qy	472	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	491	
Db	1104	GTACATCAGCGGCACC-----GCCGAACCGGGCGCCACCGTGACCTGACCGAT	1057	
Qy	492	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	511	
Db	1056	GGCAGCGGGAATCCGATCGGGCAG-----GTCACCGCGCAGCGGC	1018	
Qy	512	GlyLysAsnTrp-----	516	
Db	1017	AGCGGCAACTGGAGCTTCACCCCGTCCACCGCGCTGGCGGATGGAAACCGTGTCAACGCC	958	
Qy	517	GlnGlySerGluProGlyGlyValThrThrGlyGly-----	532	
Db	957	ACCGCTACGACCCGCGCGC---AATACCGCGCGGCAGGGCAGCACTACCGTGGACGCC	901	
Qy	533	SerAla-----	543	
Db	900	ATCGCGCGCGCCACGCGCACCTGAGCAATGGCAGCAGC--CTCAGCGGCAC	844	

544	Qy	GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn-	558
		:::	
843	Db	GCGGAAACCGGGCAGCAGCGGTGATCTCCACGACGCAACGCGAATCCGATCGCCGAGGTC	784
		:::	
559	Qy	-----SerTrp-	560
		:::	
783	Db	ACGCCGACGGCAGCGCACTGACCTACACCCCGTCCACCGCGATCGCCACCGGTACT	724
		:::	
561	Qy	-----AlaAsnSerGlnGlyValProAlaAsnAlaGln	571
		:::	
723	Db	GTGGTCAACGTGGTGGCGAGGACCGCGCGGTAAACAGACGCCCGCGGACGCGTGACC	664
		:::	
572	Qy	IleArgSerAsparg-----ValAsnProLysThrPheTyrAlaLeuSer	586
		:::	
663	Db	GTGATTCAGCGCGCGCGCGCGCGCGGTGATCAACCCG-----AGC	622
		:::	
587	Qy	AsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGly	606
		:::	
621	Db	AACGGCGTC-----GTCATCAGCGCG	601
		:::	
607	Qy	LeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp	626
		:::	
600	Db	ACGCCGAGCGCGGTGCACCGTGACCTCACCAGTCCGCGCGCAACCCGATAGGGCAG	541
		:::	
627	Qy	LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSer	646
		:::	
540	Db	GTC-----ACCGCGAGCGGACGGCACTGGAGC	511
		:::	
647	Qy	AlaIleThrGlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAla	663
		:::	
510	Db	TTCAACGGGGCAGCGCGCGGCAACCGGCAGGTGATCGTCGCCACGGCCACCGACCGC	451
		:::	
664	Qy	ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla	683
		:::	
450	Db	ACCGGCAATACCGCGCGCGAGCGGCCACCCAC-GGTGGACGCGGTGGCGCGCGCGCC	392
		:::	
684	Qy	TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyr	703
		:::	
391	Db	-----GGTGATCGATCCGAGCAACGGCACGACCATCAGCGGCACCGC	350
		:::	
704	Qy	GlyAsnTrpGlyGlnAlaIleThrGlyAsp-----HisAlaAsnLeuArgArg	719
		:::	
349	Db	GGAGGCGCGGGCGCAA-----GGTGATCCTCACCAGCAGGCAACGGCAA	308
		:::	

## RESULT 4

```

RESOLUTION 13773
US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;

```

Alignment Scores:		
Pred. No.:	2.92e-08	Length:
Score:	201.50	Matches:
Percent Similarity:	31.62%	Conservative:
Best Local Similarity:	22.10%	Mismatches:
Query Match:	4.99%	Indels:
DB:	4	Gaps:
		43
		4131

US-09-917-376-3 (1-740) x US-09-252-991A-13773 (1-4131)

QY	48	AsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr-----	65
Db	3954	AACGGCAACCGATTGGCCAGGTCAACCGCGAGCGGTAAGTGGTCTTCACTCCA	3895
QY	66	-----AsnGlyValVal-----SerIleAlaAlaAspProIleAsnThr	78
Db	3894	GGCATCCCGTCCGATGGCAAGTGGTCAACGGTGGCGCGAGCCCAAGCAATGTC	3835
QY	79	AsnLysValTrpAla-----AlaValGlyMetTyrThrAsnSerTrpAsp	93
Db	3834	GACAGTGGCGCGGTGTACTGTGGATGGCGTGGCCCCGGCGCGCGGTGATCGAT	3775
QY	94	ProAsnAspGlyAla---IleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr	112
Db	3774	CCGAGCAACGGCACCCAGAGATAAGCGGTACCGGAGGCGCGGACGGTGTCTCTCAC	3715
QY	113	ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla	132
Db	3714	GAT-----GGCGGCGGCAACCCG-----ATCGGCCAGCGCACCCGCC	3679
QY	133	ValAspProAsnAsnAspAsnIleTyrPheGlyAlaProSerGlyLysGlyLeuTrp	152
Db	3678	-----GCGGAGCGCGCACTGG	3661
QY	153	ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr	172
Db	3660	ACGTTCAACCCG-----GGCACC	3643
QY	173	TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnValValTrp	192
Db	3642	CGCTGGCCCAACGGCACCTGTATCAACGGCC-----	3613
QY	193	ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePhe-----	210
Db	3612	GTGGCCAGGACCCGCGCGCAATACAGCGGTCCGGCCAGCGTCAACGTCGATGCCATC	3553
QY	211	-----ValGlyValAlaAspProAsnProValPheTrpSer-----	223
Db	3552	GCCCGCGCGCGCGGTGATCAATCCGAGCAACGGAGTGTCTATCAGCGGTACGGCGGAA	3493
QY	224	-----Arg	224
Db	3492	GCCGGGGCCACGGTGATCTCACCGAGCGCAACCGCATCGGCCAGGTCAACGCC	3433
QY	225	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro-----	237
Db	3432	GACGGCAGCGCAAGTGGCTTTCAGCCCGCCACCGCGTGGCCAAATGGCAGGTGATC	3373
QY	238	-----Thr	238
Db	3372	AATGCGTGGCCGAGGACCGCCCGGCAACACAGCAGTCCCAACGCGCCACCGTCTGAC	3313
QY	239	GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr	258
Db	3312	TCGTGGCCCAACGACCCCGGTGATCTGATCCAGCAACGGTAGGTGATCCCGGTACC	3253
QY	259	SerAsnThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpLysPhe	275
Db	3252	GCCGAGCTGGTGCCACGGTGATCTCACCGAGCGCAACCGCATCGCCAGGTC	3193
QY	276	SerValThr---SerGlyThrTrpThrArgIleSerProValPro-----	289
Db	3192	ACCGCGGATGGCAGCGCAACTGGAGATTTCACGCCCGGCAACCGCGGTCTTCAATGCAAG	3133
QY	290	-----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr	304
Db	3132	GTGGTCAATGGTGCCCAAGGACGCTGCGCGCAACACCGAGCGGCCCGCGCAGCACCG	3073
QY	305	IleAspArgGlnHisPro-----AsnThrIleMetValAla	316



3072	GTGACTCGGTGGCGCGCGCGCGCGGTGATCGACCCGAGCAACGCGCAGCTGATCGCC	3011
317	ThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly----	334
3012	GGTACCGCGGAGCGGTGCGACGGTGATCTC-----ACCGATGCGCGCGGCAAC	2962
335	-----AlaThrTrpThrArgIleTrpAspTrpThrSerTyrrProAsnArgSer	350
2961	CCGATCGCCAGCGCCACCGCGATGGCAGCGGCAACTGAGACTTCACTCCCGCGGCA	2902
351	LeuArgTyrValLeuAspIleSerAla-----Glu	360
2901	CTGCGCCACGCGACGGTGATCAATGGCGGTGGCCCGCAGGATCGCGCGCGCAATAC	2842
361	ProTrpIleuThrPhe-----GlyValGlnProAsnProValProSerProLysLeu	378
2841	CCGACCAAGCACCGGTGGAGCGGTGGCCCGCCACCCCGGTGGTCAACCCGAGCAAC	2782
379	GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly	398
2781	GGC-----AGCGTGATCGCGGT	2764
399	Thr-----GlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleThrIysTrpAsnSerGly	415
2763	ACCGCGGACCGCGGCCACCGGTGATCTCTCACC-----GACGGCGGC	2722
416	GlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeu	435
2721	GGCAAC-----CCGATCGCGCAGGTCAACCGCCAGCGCAGCGCACTGGAGC	2674
436	IleSerProSerGlyAlaProLeu-----IleSerAlaLeuGly----	449
2673	TTACAGGCC-----GGCACGCGCTGGCCACCGGCTCGGTGATCAATGCGTGGCCAC	2620
450	AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaValProSerThrIlePheThr	468
2619	GACCCCGCGCAACACCAACCGCGCGCGCCAGCACCAACCGGTGGACTCGGTAGCCCG	2560
469	SerProValPhe-----ThrThrGlyThrSerValAsp---TyrAlaGluLeuAsnPro	485
2559	ACCCCGGTGCTCGATCCGAGCAACGGTACCGGTGATCAGCGGTACCGCCGAAGCCGGG	2500
486	SerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisVal	505
2499	ACGGTGATCTCACCGACGCGCGCGGCAACCCGATACGGCAGGCC-----	2455
506	AlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThr	525
2454	-----ACCGCCGATGGCAGCGCAACTGG-----AGCTTCACTCCGGGCACACCG	2407
526	ThrGlyGlyVal-----AlaAlaSerAlaAspGlySerArgPheValTrpAlaPro	543
2406	ACCAACGGCAGGTGATCAATGCGGTGGCGCAGACCGCCCGGCAACACAGCGGTCCG	2347
543	-----	543
2346	GTCAGCACACAGTGGAGCGGTGGCCCGCGCCACCCCGGTGATCGACCCGAGCAATGGT	2287
544	-----GlyAspProGlyGlnProValValTyrAlaValGlyPheGly	557
2286	GTCAAACTCAGCGGCACCGCGCAACCCGGGTCCGGGTGATCTCTCAACCGATGGCAAT	2227
558	Asn-----SerTrpAlaAlaSerGlnGlyVal	566
2226	AACCCGATCGGCCAGACCTCGCCGACGGTACGCGCAACTGGACTTCAACACCGGGCAG	2167
567	ProAlaAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSer	586
2166	CCG-----CTGGCC	2158
587	AsnGlyThrPheTyrArg-----SerThrAspGly-----	596
2157	AAACCGCAGGTGTCAACGCGCGTGGCCGACGACCCCGCGCGCAATACAGCGGTCCG	2098

## RESULT 5

US-09-103-840A-1/c

US-03-103-840A-1/C  
: Sequence 1. Application US/09103840A; Sequence 1, Application  
: Patent No. 6294328; PACEIL NO. 6294328  
; GENERAL INFORMATION:

APPLICANT: FLEISCHMAN. Robert D.

APPLICANT: FLEISCHMAN, ROBERT  
APPLICANT: WHITE, OWEN R

APPLICANT: FRASER, OWEN R.  
APPLICANT: FRASER, CLAIRE M.

1. APPLICANT: FRASER, Claire M.  
2. APPLICANT: VENTER, John C.

APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

[illegible]; TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 242,000,000,000

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/0

**; CURRENT FILING DATE: 1998-**

; NUMBER OF SEQ ID

; SOFTWARE: F

; SEQ ID NO 1

; LENGTH: 4411529

```

;
; TYPE: DNA

```

Alignment scores:

Alignment scores:

Pred. No.:	0.0104	Length:	4411529
Score:	192.00	Matches:	194

Score:	192.00	Matches:
Percent:	25.04%	Consensus:

Percent Similarity:	35.04%	Conserva
Post Local Similarity:	33.50%	Mismatch

Best Local Similarity: 23.60%

Query Match: 4.76% Indels: 5

DB: 3  
Gaps: 3

[illegible]

US-09-917-376-3 (1-740) x US-09-103-840A-1 (1-4-

Qy 10 AsnValAlaIleGlyGlyGlyGlyPheValAsp




US-09-917-376-3 (1-740) x US-09-103-840A-1 (1-4411529)

10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly----- 27

QY  
IU ASHVAIAIIEGLYGTYGYFNEVATASBPGRYTEVAIFNEASINIGLY-----Z/  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qv 28 -----A]a proG]yT]e]e]uTyrValArgThrAspT]eG]v-----G]vMet 41

QY 28 -----AlaPFGGLYLLLeuLeuYFvaIaTARGInHApRLeGLY-----GlyMet 41

42 TyrArgTrpAspAlaAlaAsn--GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60

QY 42 YUAGIGIPASPAIAATAASH--GYUAGIGIPRIIEFIOLEULEUASPIRPAVAGIYIP 00

61 AsnAsnTrpGlvTvrAsnGlvVal--ValSerIleAlaAlaApp----- 74

QY  
01 ABITASIIIPRGUYUASINGUYVAI--VALZELLEATATAABP

75 ---proIleAsnThrAsnIysVa[TrpA]aVa[GlyMetTyrThrAsnSerTrpAsp 93

QY /3 ---PFOLEASHNINIGASNLVSVALIRPRAIAATAVALGTYMECTYUINLASNIZETIRPASR 53

```
QY 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113
Db 427012 AACAACTTTGGC-----
QY 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyCyluArgLeuAlaVal 133
Db 427000 -----ATGGCAACGGGGCAACTTCAACGACGGGCATTGCC 426965
QY 134 AspProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLys---GlyLeuTrp 152
Db 426964 AACCGGCAACAACAACTCGCTGTTCAACACGGGCAACAACACGTCGGCATCTGG 426905
QY 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172
Db 426904 CTGACGGCGGACGGCTGTCTCGGCTTCAGTCCCTGAACTCGCGCGCGGCAACACCGGT 426845
QY 173 TyrlleAlaAsnProThrAspThrThrGlyTyTrpGlnSerAspIleGlnGlyValValTrp 192
Db 426844 TTCTTCAACTCGGCACCGCCAAACCCGGC----- 426815
QY 193 ValAlaPheAspLysSerSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211
Db 426814 ---TTGTTCAACTCGGCACCGCAACACCGGCTTGTTCAACTCGGCACCGGCAACGTC 426758
QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231
Db 426757 GGCATCGGCAACATGGGCAACCGCGGCTTCGGGCTCGGCTATCCGGCGACGACCGAGTG 426698
QY 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251
Db 426697 GGCATCGGCGGCAACCACTCGGCGAGTTTC---AACTCGGCTGTGTTAACTCGGCGACC 426641
QY 252 HisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyTrpAspGlySerGly-As 271
Db 426640 GGCATGTGGCATCGGCAACTCGGCGACCGGCAAGCTCGGCATCGGCAACCGGCAACC 426581
QY 271 pValTrpLysPheSerValThrSerGlyThr-----Tr 282
Db 426580 GGCAACACCGGCATCGGAACAGCGGCAACTACAAACCGGCTGTGTTCAACGGGGCGCTG 426521
QY 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296
Db 426520 GTCAACACCGGCATCGCAACCGGCGCAACCAACACCGGCTGTGTTCAACATCGGCACC 426461
QY 296 pTyPheGlyTyTrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316
Db 426460 TTCAACACCGGCATCGCAACCGGCGCACTACAAACCGGCTCTTCAACACCGGTAGC 426401
QY 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331
Db 426400 TACAACACCGGCATCGCAACCGGCGAGACTAGCGGCAACCGGCGGTTCATCACCAGCGCAGC 426341
QY 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer-- 345
Db 426340 ATGAACAACCGCTGTCTCTGGCGCGCGGCGGCGGCTGTCTGGCGGCGCAACTACACC 426281
QY 346 -----TyTrpAsnArgSerSerLeuArgTyTrpValLeuAspIleSerAlaGl 360
Db 426280 ATCACCATCGAGGACCTCGCGGCTTCTCAATGTGACATCCCGGTCAACATCCCATC 426221
QY 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380
Db 426220 ACCG---GGACATCAACCAATGTCTCAATCCCGCCATACGTTCCCCAGA-ATC----- 426170
QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyTrpGlyTh 399
Db 426169 ----GACGCGAGGGAAGGCTCGACATAGGATCTCTCAGTGGCAGCGCTTGTGGCCCGGT 426114
QY 399 rGly---AlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418
Db 426113 CGGTCCGATCACCCCTGATGGGGGACGCGTTCGGCGCCCGCTGGACACACCCATCGAAAT 426054
```

RESULT 6

```

US-09-103-840A-2/c
; Sequence 2, Alignment US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 0.0564 Length: 4403765
Score: 184.00 Matches: 204
Percent Similarity: 32.65% Conservative: 116
Best Local Similarity: 20.82% Mismatches: 306
Query Match: 4.56% Indels: 361
DB: 3 Gaps: 50

US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)
Qy 14 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu--- 32
Db 3745039 GCGGTGGCGGTCAACCGGATGCG-----GATCGGGCGGTGGCGGTGCGAATCT 3744989
Qy 33 -----TyrValArgThrAspIleGly 39
Db 3744988 GTTGGGCGAGAACGCCGACGATCGCGCGCGTCGAGCGCGAGTACGAGTTGATGTGGGC 3744929
Qy 39 ----- 39
Db 3744928 CGCGATGTGGCGGCGATGCGCGCTACCATTCGCGCGGTGCTGCTCCCGGGCGGT 3744869
Qy 40 -----GlyMetTyrArgTrrAspAlaAlaAsnGlyArg----- 50
Db 3744868 GCGGGCGTTCAGCCACCGCGCAGG-CGTGGGGGTGTGTGCGCGCTTCTCAATG 3744810
Qy 51 -----Trp-----IleProLeu 54
Db 3744809 CTCTATTGCGGACCGCGAGATGTTAGGCTTAACCGCGGCTTTGGCAATGTCGGTA 3744750
Qy 55 LeuAspTrrValGlyTrrAsnAsnTrrGlyTrrAsnGlyValValSerIleAlaAlaAsp 74
Db 3744749 ATTACACGTGCGGTTGGGCAATGTCGGATAT-----TCAACCTGGGCGCAG 3744702
Qy 75 ProIleAsnThrAsnLysValTrr-----AlaAlaValGlyMetTyr 88
Db 3744701 CCAATGTCGGTGCAGAAATTTGGTGCTGCCAACGCCGTAGCGGAATTCGGTTTCG 3744642
Qy 89 ThrAsnSerTrrAspProAsnAspGlyAlaIleLeuArg----- 101
Db 3744641 GCAATATCGCAACCGCAACTTCGGGTTCGGCAACTTCGGGTCTTGGGTTCGCCCGCGGCA 3744582
Qy 102 -----SerSerAspGlnGlyAlaThr 108
Db 3744581 TGGCAATATTGGTTGGGCAATTCGGGCGACGACGAACTACGCGCTTCGAAACCTGGGTG 3744522
Qy 109 TrrGlnIleThrProLeuProPheLysLeu-----Gly 119
Db 119 -----

```

QY 361 -----ProTrpLeuThrPheGlyValGlnProAsnProProValPro 374  
Db 3743424 GCCAGCGTCACTGTCGGCCTATCTCTGATCAACGGCGTT-----AATATCCGGCG--- 3743374  
QY 375 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer---Asp 393  
Db 3743373 ACCCGGGTTTGGCAACACGACCGCTCGCTCGCGGGTTTCTTCAACTCCGGCGAC 3743314  
QY 394 ArgMetLeu-----TyrGlyThrGlyAlaThrLeuTyr-----Ala 405  
Db 3743313 GTTGGGTGTCGGGCTTCGGGAATTCGGTTCGGGCGAGCTCGGTTGGTGAACACAGGCG 37433254  
QY 406 ThrAsnAspLeuThrLysTrpAspSergly-----Gly 416  
Db 3743253 CAGACCGAGTGTGGCGGGTTCGGGTTCGCCAATTCGGTTCGGTTCGGTTCGGGT 3743194  
QY 417 GlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIle 436  
Db 3743193 GTGCTGAACCTCGGCTCGGGTGTGTCGGGCTGTACAACACCGCGCG--- 3743146  
QY 437 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456  
Db 3743145 TTGCGCGCGGGACCGCGGGTGTCTCGGGCATCGGCAATGTTGGTGAG----- 3743095  
QY 457 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 476  
Db 3743094 -----CAGCTGTCTGGGGTGTCTCTCGGGCGGACG 3743065  
QY 477 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 496  
Db 3743064 GCA-----CTCAACACGAGCCTCATCATCAATCTCGGGTTGGCCGATGTG 3743020  
QY 497 SerSerGlnProAsnAspArgHisValAlaPhe-SerThrAspGlyGlyLysAsnTrpPh 516  
Db 3743019 GCGAGCGTA-----AACGTCTGGTTTCGGCAACGTCGGGACTTCAACCTGGGT 3742972  
QY 516 eGln-----GlySerGluProGlyGlyValThrGly 527  
Db 3742971 GCGGCCAATATCGCGACTTGAACGTGGTTCGGCAATTCGGCGCGCGCAACGTGGG 3742912  
QY 527 yGlyThrValAlaAla-----SerAlaAspGlySerArgPheValTrpAlaProGlyAs 545  
Db 3742911 TTGCGCAATATCGCGACTTCACTTCGGGTTCGGCAATTCGGGCTCGCGCGCGGCGCTG 3742852  
QY 545 pProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer---G1 564  
Db 3742851 GCGGGG-----TGGCAACATCGGGTTG 3742828  
QY 564 nGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAl 584  
Db 3742827 GGCAATGCGCGCAGCGGCA-----AC 3742807  
QY 584 aleuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAl 604  
Db 3742806 GTCGGTTCGGCAACATCGGTGTGGCAACATCGGTTCGGTAACACCGCGCACCAAC 3742747  
QY 604 aAlaGlyLeu-----ProSerSe 610  
Db 3742746 CTCGGGATGGGCTGACCGGGGCAACACGAGACTGGGATCGCGGCTTGAACCTCGGT-GC 3742688  
QY 610 rGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAl 630  
Db 3742687 CGGCAACATCGG---TTGTTCAACTCCGCGCACCGCAACGTCTGG---TTGTTCAACTC 3742634  
QY 630 aAlaSerSer-----GlyLeuTyrHisSer-----ThrAsnGlyGlySerSe 644  
Db 3742633 CGGGACCGGGAACCTCGGGTGTGTTCAACTCGGCGAGCTTCAACACCGCGCATCGGCAATGG 3742574  
QY 644 rTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaPr 664  
Db 3742573 CGGAACCGGCGAGTACTCGGGCTTTTCAATCGCGTAATTTCAATACCGGTGTGGCCCAACCC 3742514

QY 664 oGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyr 684  
Db 3742513 TGGGTCTGATACACACGCGGCTTCAATGTGGGTGACACCAAC-----ACCGTGGTTT 3742460  
QY 684 rArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrG1 704  
Db 3742459 CAACCCGGGAGCATCAACACCGGCTGTCTCAACACCGGCAACGC-CAACACCGGCGTGG 3742401  
QY 704 yAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGly 723  
Db 3742400 CCAATTCGGGCA-----TGTGACACCGCGCCCTCATGTCTGGG 3742361  
RESULT 7  
US-09-328-352-1377  
; Sequence 1377, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1377  
; LENGTH: 11679  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1377  
Alignment Scores:  
Pred. No.: 1e-05 Length: 11679  
Score: 181.50 Matches: 204  
Percent Similarity: 31.05% Conservative: 90  
Best Local Similarity: 21.54% Mismatches: 344  
Query Match: 4.50% Indels: 309  
DB: 4 Gaps: 54  
US-09-917-376-3 (1-740) x US-09-328-352-1377 (1-11679)  
QY 7 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGly----- 21  
Db 1627 ACGTGGACA---GTAAGTGTGCGGGTAGTGTGTGGTGTGTGATGCGAGATAAACGATT 1683  
QY 22 -----IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArg---ThrAsp 37  
Db 1684 GATGCTAAAGTAACGTTTACAGATGACAGAGTAAACGACACTGTTAAACATACGCNA 1743  
QY 38 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 57  
Db 1744 ATT-----TATACATTAGACACAGCT--GTCCTCGACGCGCCAGTAAATCGACCCA 1791  
QY 58 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyVal-----ValSerIleAlaAlaAspPro 75  
Db 1792 GTT-----AACGGGACAGACCCCAATTACAGAGTACAGCAGAACCT 1830  
QY 76 IleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsn----- 90  
Db 1831 GGTTCACAGTAACAGTA-----ACCTATCTCTATGGTGCACACGCAACA 1875  
QY 91 -----SerTrpAsp---ProAsnAspGlyAlaIleLeuArgSer 102  
Db 1876 GTTGTAGCAGGACCGGAGTGTGTGTCAGTACCAACCCCTGGC-----CTTAATGAT 1929  
QY 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn--- 121  
Db 1930 GGTGACGAAGTTGAGGCAATTTGCTTACAGATCCA-----GCAGGCAACCCA 1974  
QY 122 ---MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn----- 137  
Db 1975 TCTTTGCCAGGTACAGCTACTGTTGAT-----GCAGTTGGTCCAAATACCGATCGTGT 2028  
QY 138 -----AspAsnIleLeuTyrPheGlyAlaProSerGly 148

Db 2029 AACTTTACGGTTGATTACAGTACCAATGTGATTAAATGATCATCAGAACCGTCAGGC 2088  
|||||  
Qy 149 Lys-----GlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
|||||  
Db 2089 AACGTTACTGTTACTGGTGTATTGAAACAGTTCGGCAGATGCAGCAATACAGTGGTC 2148  
|||||  
Qy 165 ThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
|||||  
Db 2149 ACTGTTGTAATCAATGCGCAGACGTATACCTGCACTGTAGATAGACACAGGC----- 2202  
Qy 185 SerAspIleGlnValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 204  
|||||  
Db 2203 -----ACATGGACAGTAACGCTACCGAGTACCTAACTACCTCGGAT 2244  
Qy 205 AlaSerLysThrIle----- 209  
|||||  
Db 2245 GCAGATAGACGATTGATGCTAAAGTAACGTTTACAGATGCGGAGGTAATAGCAGCAGT 2304  
Qy 209 ----- 209  
Db 2305 GTTAACGATACAAACATATACATCGATACCACTGCACCTGTACCGACCACTAATTAAC 2364  
Qy 210 -----PheValGlyValAlaAspProAsnAsnProValPheTrp 222  
|||||  
Db 2365 CCGGTTAAACGGGACACACCGGATTACAGGTACGGCAGACCTGGTTCAACAGTAACCTGTG 2424  
Qy 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro-----ThrGlyPheIle 241  
|||||  
Db 2425 ACTTATCCAGATGGCAGTACACACAGTTGTGTGACGGACCGGATGGCAGTGGACAGTA 2484  
Qy 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261  
|||||  
Db 2485 CCAACCCAGGTTTAAATGATGC-----GATAAGTTACAGCAATTTGCTACAGATCCAGCA 2541  
Qy 262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281  
|||||  
Db 2542 GGCACACCA-----TCATTACCAAGTACA 2565  
Qy 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
|||||  
Db 2566 GTACTGTTGATGCGATGGTCCNAATACCGATGGTGTAAAT-----TTACGGTTGAT 2619  
Qy 302 GlyLeuThrIleAspArg----- 307  
Db 2620 TCAGTAACAGCTGACAATGTGATTATGTCATCAGAACGATCAGGCAACGTTACTGTTACA 2679  
Qy 308 -----GlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 2680 GGTGTATTGAAAAACGTTCCGGCAGATGCAGCAATATACAGTG---GTCACTGTTGTGATC 2736  
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
|||||  
Db 2737 AATGCCACAGCTATACCTGCACTGTAGATAGACACAGCAGGC-----ACATGGACA--- 2787  
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValIleAspIleSerAla 359  
|||||  
Db 2788 -----GTAAGCTACAGTAGTACTGACTGCGGATGACAGTAAGACGATT 2835  
Qy 360 GluProTrpLeuThrPhe----- 365  
Db 2836 GATGCCAAGTAACTGTTTACAGATGCAGCAGGTAAACAGCAGTGTAAACGATACACAC 2895  
Qy 366 -----GlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380  
Db 2896 ACATATACAGTTGATACGGTTTGCACCAATGCACCGGTG----- 2934  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAsp----- 393  
Db 2935 CITGATCCGATCAATGCACACACCGGATGACGGGTACGGCAGACGCTGGTTCAACAGTG 2994  
Qy 394 ArgMetLeuTyr-----GlyThrGlyAlaThrLeuTyrAlaThrAsnAsp----- 408  
|||||

Db 2995 ACTGTGACTTATCTCTGATGGCACCACTGCAACAGTGGTAGCAGACCGGATGGTAGTGG 3054  
Qy 409 -----LeuThrLysTrpAsp 413  
Db 3055 TCAGTACCAAAACCCAGGTAACTGGTGGATGGCGATACAGTGTCAACAGCAACTGAC 3114  
Qy 414 SerGlyGlyGlnIleHisIle-----AlaPro 422  
Db 3115 CCTGCGAGCAACACTTCATTGCGCAGGTACAGGCACAGTTCCTCAGCAGACATCACAGCACT 3174  
Qy 423 MetValLysGlyGluGluThrAlaValAsnAsp-----Leu 435  
Db 3175 GTGGTT---CGCTGGATGACGTGTTCAGCAATACAGACACACAGCATTACCGGTACA 3231  
Qy 436 IleSerProProSerGlyAlaProLeuIleSerAla----- 447  
Db 3232 GTGACGATCCGACAGCCACTGTAGTTGTCAATGTGGATGGCTGTGACTATCCGGCAGTG 3291  
Qy 448 ---LeuGlyAspLeuGlyGlyPheThrHisAlaAsp-----ValThrAla 461  
Db 3292 AACAAATGGTGAC---GGCACCTGGACGCTTGCAGACAATACACTTCCTCGCTTAACCTGAT 3348  
Qy 462 ValProSerThrIlePheThrSerProValPheThrThrGly----- 475  
Db 3349 GTTCCACACACCACTTACCTGACTGCAACAGATGCAGACGCAATGCAGGTACAGATACA 3408  
Qy 476 -----ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493  
Db 3409 GCAGTGGTCAGGATGATACACAGCACCAGCAATGCACCGGTACTTGTATCCGATCAATGCG 3468  
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513  
Db 3469 ACTGACCCA-----GTGAGCGGTACAGCAGAACTGGTTCA 3504  
Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSer 533  
Db 3505 ACC---GTGACTGTGACTTATCTCTGATGGCACCACT-----GCAACAGTGTAGCAGGC 3555  
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProVal----- 550  
Db 3556 ACAGATGGTAGC-----TGTGATGACCAACCCAGGTAACTGGTAGATGGTGTAT 3606  
Qy 551 ---ValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569  
Db 3607 ACAGTAACGTCAACA-----GCAACTGACCTGCAGGCAATACATTCATTCGCGAGTACA 3660  
Qy 570 AlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589  
Db 3661 GGCACAGTCTCTGCAGACATCACAGCACCT-----GTGGTTGCGCTGGATGAGCTGTG 3714  
Qy 590 PheTyrArgSerThrAspGly-----GlyValThrPheGlnProValAlaAlaGlyLeu 607  
Db 3715 ACCAATGACAGCACACACGACTTACGGGTACAGTGAACGATCCGACAGCCACTGTAGTG 3774  
Qy 608 ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeu 627  
Db 3775 GTGAAT-----GTAGATGGCACTGACTATCCGGCAGTGAACATGGCGAGGC---ACC 3825  
Qy 628 TrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla 647  
Db 3826 TGGACGCTTGCAGACAATACACTTCCAGTGTGGCAGACGCTCCACACACC---ATTACC 3882  
Qy 648 IleThrGlyValSerSerAlaValAsnValGly-----PheGly 660  
Db 3883 GTGACTGCAACAGATGACGACGCAATGCGAGGTACAGATACAGCAGTGGTGAACGATTGAT 3942  
Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyVal 680  
Db 3943 ACCACAGCAACCAATGCA-----CCGGTACTTGAATCAATGATCGACTGACCCAGTG 3996  
Qy 681 ThrGlyAlaTyrArgSerAspAspCys-----GlyThrThrTrp 693  
Db 3997 AGCGGTACAGCAGAACTGGTTCAACCGTGACTGTGACTTATCTCTGATGGCACCCTGCA 4056



QY	317	-----	317	QY	589	ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro	608
Db	19531	GGCAGATATGTTGAAGTGGTATCAACGGCAAAACCTACACCTCTGAACCGGGCGCGCG	19590	Db	20392	ACGTTTGTCTGGGGGAGAACACACCAATATCTTCTGTAGCGGGCGGAGTGGCGCT	20451
QY	318	-----GlnIleSerTrpTrp-----ProAspThr-----	325	QY	609	SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys---GluGlyAspLeu	627
Db	19591	GTAGTGGTTCGATCCGGGCGACAAACACCTGGTATGTACAGTTCCGGATACCGATCGCTG	19650	Db	20452	ATGACGACGAGCGTTTCCATGACCTGGCGGATTCGATGGTGGCGGATATGATCTC	20511
QY	325	-----	325	QY	628	TrpLeuAlaAlaSerSerGly-----LeuTyrHisSerThrAsnGly	641
Db	19651	ACAGTTTCGGCGACCGCTATACCGTTACTGCGCAGGTAAAGTTCCGGCGGTAAACGGC	19710	Db	20512	TTCTCTCCCGCCAGCCAGCGAGAGCTAACTACGGCTCGCTGTTATTCAACACCAACGGC	20571
QY	326	-----IleIlePheArgSerThrAspGlyGlyAla	335	QY	641	-----	641
Db	19711	AATAAGGCCAATATTAGCAACGGCAGCGGTACGGTTAACCGCGGATGATTACACACCG	19770	Db	20572	GTACTGGTTGCCCGGTGGCGGTGGCGCAACCGCAACACCTACGCCAGCCAGTTTAGC	20631
QY	336	ThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeu	355	QY	642	-----	Gly 642
Db	19771	ACCTGGACTACCGCCAGCAAAACACCGCCCTGG-----GGGCTGACCTACCGCCTC	19821	Db	20632	CTGGCGGTGAGTGAACACGACGCGGCTGATGGATATCGCCCGTATCGCCAGACCGGG	20691
QY	356	AspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSer	375	QY	643	SerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer	662
Db	19822	GAC---TCGACGGGATGTGACGGTGTGGCAACACGAGGTAAATGCAATCGACTGAC	19878	Db	20692	CAGTCGTAT---CTTTATCTAATACGTACGCAACGCCAGCAAC-----TGCACCAATCG	20742
QY	376	ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMet	395	QY	663	AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr---	681
Db	19879	CCA---CTCACTGTCTGAGACCGCGCTGAC-----	19908	Db	20743	GCCTCGCGCGCAGCCAGCAGC-----GGTACCACGCGGGGTGGCGCA	20787
QY	396	LeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGly	415	QY	682	-----GlyAlaTyrArg-----SerAspAspCysGlyThrThr	692
Db	19909	CTGTATCAGCGCGCAACAAAC---TACGCCACAGCTCCATTGCCGATTACGACCGTAAC	19965	Db	20788	ATGCACTACGACTGGGACGCGCGGTGGATGTGCTGTGTCTCCAAACAGTCGGGAGCGTG	20847
QY	416	GlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeu	435	QY	693	TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGly	712
Db	19966	GGC-----	19968	Db	20848	TTCTCGCGCGCAACACCAACGACGCGTACGCTACGCG-----ACTTCG	20889
QY	436	IleSerProProSerGlyAlaProLeuIleSer-----AlaLeuGlyAspLeu	451	QY	713	AspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGly	732
Db	19969	-----ACGGCGATCTGTTTATCACCGGTGATGACTACGTTACGGGCTATATT	20016	Db	20890	CTACACTCGGCATC-----ACCGATCCCAACGCGATTAAAGTCTATTACGGC	20937
QY	452	GlyGlyPheThr---HisAlaAspValThrAlaValProSerThrIlePheThrSerPro	470	QY	733	Asp 733	
Db	20017	AACGGCTTTACCAATAACCGCGATGCGACC-----TTTTCACGCGCT	20058	Db	20938	AAT 20940	
QY	471	Val---PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleVal	489	RESULT 9			
Db	20059	ATTTCAGGTCACCGTCGCGCACCTGACGTGGTACGGC-----TCGATTGTG---	20103	US-09-103-840A-2			
QY	490	ArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThr	509	/ Sequence 2, Application US/09103840A			
Db	20104	-----GCATTGTATTAAGAGGGC-----	20121	/ Patent No. 6294328			
QY	510	AspGlyGlyLysAsnTrpPheGlnGlySerGluProGly-----	522	/ GENERAL INFORMATION:			
Db	20122	GACGGCTATCTCGACTTCGATTGTTGATGCGTGGCGGCGCGGACTCCAAACACCTTCCTG	20181	/ APPLICANT: FLEISCHMAN, Robert D.			
QY	523	-----GlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg	538	/ APPLICANT: WHITE, Owen R.			
Db	20182	TGGAACAAACGACGCGCGTGTAGGCACTCCACCGCTCGAAGACGCGCGGTAGCGCC	20241	/ APPLICANT: FRASER, Claire M.			
QY	539	PheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsn	558	/ APPLICANT: VENTER, John C.			
Db	20242	ACGGTGGCGCGG-----GCGGTGACGGGGTATCTT	20271	/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM			
QY	559	SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal---	577	/ TITLE OF INVENTION: TUBERCULOSIS			
Db	20272	TCGCTCAACGAGGTCTGGCGCTCGATCTGCAACATGACGCGGAGGATCGACCTGGTTACG	20331	/ FILE REFERENCE: 24366-20007.00			
QY	578	-----AsnProLysThrPheTyrAlaLeuSer-----	588	/ CURRENT APPLICATION NUMBER: US/09/103,840A			
Db	20332	CACACCTATAACCTGAACAACTATTACAGCTGTCTTCGCTCATCAACCGGGGAATGGG	20391	/ NUMBER OF SEQ IDS: 2			
				/ SOFTWARE: PatentIn Ver. 2.1			
				/ SEQ ID NO 2			
				/ LENGTH: 4403765			
				/ TYPE: DNA			
				/ ORGANISM: Mycobacterium tuberculosis			
				/ FEATURE:			
				/ OTHER INFORMATION: CDC 1551			
				/ OTHER INFORMATION: "n" bases at various positions throughout the sequence			
				/ OTHER INFORMATION: represent a, t, c or g			
				US-09-103-840A-2			
				Alignment Scores:			
				Pred. No.:	1.09	Length:	4403765
				Score:	170.00	Matches:	171



	Percent Similarity:	33.42%	Conservative:	82
	Best Local Similarity:	22.59%	Mismatches:	324
	Query Match:	4.21%	Indels:	184
	DB:	3	Gaps:	32
US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)				
Qy	82 TrpAlaValGlyMetTyrThrAsnSerTrpProAsnAspGlyAlaIleLeuArg 101       ::: 			
Dd	3791642 TGGCGCCGAAGCCTGTACGCCAGGCATCAACCATTCT-----CGA 3791686       ::: 			
Qy	102 SerSerAspGlnClalathrTrpGlnIlleThrProLeuProPheLysLeuGlyGlyAsn 121 :::    :::			
Dd	3791687 GAAGCTGATCGG-----ACGACCGAGCGGTGGTGGA 3791719 :::    :::			
Qy	122 -----MetProGlyArgGlyMetGlyGluaRgLeuAlaValaspProasnAsn 137       			
Dd	3791720 CCAGGCCGTAGCCGCTGGCCGGTT--AATGCCCGACTCAAGCACGCCGTCGTGCAG 3791776       			
Qy	138 AspAsnileLeuTyrrPheGlyAlaProSerdglyLyasGlyLeuTyrrArqserThrAspsEr 157 :::    :::			
Dd	3791777 TC-TCCACGATGCCGACGCTCGCCCAT CGCAAAGGGACGATTGGCAAGCCGTCGAGT 3791835       			
Qy	158 GlyAlaThrtTpSrGlnMetThranPherProaspValGlyThryzrlalealanPro 177       :::			
Dd	3791836 TCGGCTACAAGGCCACAGTCTGCACAAACCGCGAGTGTCTATCTCTGG----- 3791883       			
Qy	178 ThrAsptHrThrglyTyrrGlnSerAsplieGInGlyVal-----ValTrpValAlaPhe 195       			
Dd	3791884 -----ACCACAGCTGAGCTCGGAACCCGCGAGATACCGCANITGGCACCCGCCA 3791937       			
Qy	196 AsPLysSerSxsrSerLeudigYlGinAla-----SerLysThrillePheValGlyVal 213 :::    :::			
Dd	3791938 TCGAACGGATCAGCCGCGCACCGGACCCCACCACCGGAGTAGCCGTGATCGGGCT 3791997       			
Qy	214 AlaAspProasnMsnProvalPheTrpSrArqspGLygly----AlaThrtTp----- 230 :::    :::			
Dd	3791998 GCGGAGACGCATCGGTFCGAAGAATGATCTCCACAGCTCGGGGTGGCAACGTGGCCATCC 3792057       			
Qy	231 ----GlnAlavalProglYAlaproThrGlyPheIleproHisLysGlyValPheaSp 248 :::    :::			
Dd	3792058 CAGCGAAGCAAMACCAGCGCCACCCGCGGCGCATTCGAACACGACGCGG----- 3792108       			
Qy	249 ProValAsnHisVallLeutyriAleAthrsenThrglytyrotyrAspglySer 268 :::    :::			
Dd	3792109 -----CATTCGCGACAAAGATCAATGCGGACCG---GATCCGAAG 3792147       			
Qy	269 SerGlyaspValTrplysPheservalThrseryThrtpThrpThrArgile----- 285 :::    :::			
Dd	3792148 GAGCGATCAACCACCTCAAGCGCAGCTAGCGGTGGAAACCGCACCCAACCTCACCGGCATCA 3792207       			
Qy	286 ---SerProvalProserThraspthralaAsnAspTyrPheGlyTy-Serglyeuthr 304 :::    :::			
Dd	3792208 CCGGCCCGCAACCTGGTGGGACACGGGTCT-----TCGCCCAACAC 3792252       			
Qy	305 IleAsparg-GlnHiProasnThrilleMet-----ValalaTh 317 :::    :::			
Dd	3792253 TCGTCAAGATCAGACCCCTGGCAGCGTGCAGACACCCCCCGCCACCGACGCCAC 3792312       			
Qy	317 rGlinlleSerTrtTpProaspThrillelePheargserThrAspglyGlyAlaThr--- 336 :::    :::			
Dd	3792313 GCAGGTTCGC---CAGCCCGCGCGCTCAATGCMAACCGCGACTTTTTTCAGGTCTTAGTA 3792369       			
Qy	337 ----TrpThrArgIleTipAspTrpThrSer----- 345 :::    :::			
Dd	3792370 TTAGTGGCCCGCGCTTGGGTCCACCGGGGCCCTCGCGCAACACCGACAGCTGATGCCG 3792429       			
Qy	346 -----TyrProasn-----ArgSerLeuArgTyrrValLeuaspileseAl 359 :::    :::			
Dd	3792430 TGATCGGCGATACCTTCGACCCATTCAAGGAGAACACCCATGCTGTTTGA-TCCGG 3792488       			
Qy	359 agluPrO-----TrpLeuThrPheGlyValcIn----- 368 :::    :::			



QY 505 alAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer---GluProGlyGlyV 524  
Db 3802171 --GCCGGCGCGCTCGCGGCAACCGCGGTGTGTACGCGCAACCGCGCGCGCGCA 3802228  
QY 524 alThrThrGlyGlyThr-----ValAlaAlaSerAlaAspGlySerA 538  
Db 3802229 TCGGGGGCGCGGACCGGAAACCGGTGTACCGCGGGCGCGCGCGCGCGCGG 3802288  
QY 538 rgPheValTrp-AlaProGlyAspProGlyGlnProValValTrpAlaValGly--Phe 557  
Db 3802289 CCTGGCTGTGGGCACCGCGCGGGCGCGCGGACCGCGGTGTGTGTGTGTGTGT 3802348  
QY 557 lyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSerAspArgV 577  
Db 3802349 GCGACGGCGGGCGCGGCGCACCGCGGCAACCGCGGCGCGGCTTTAAACAGC----- 3802401  
QY 577 alAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyG 597  
Db 3802402 -----TTGACCTCTTTCGTGTGGCG 3802420  
QY 597 lyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetP 617  
Db 3802421 GCGCGCGCGCGCGGTGGCGACCGCGGCTGTTCGGCGCGCGCGGACCGCGGCGGAC- 3802479  
QY 617 heHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyxH 637  
Db 3802480 --GGCGGCATCGCGGCGGCAAAACACGAGACCGCGCGCGCGCGCGCGCGCGCG 3802537  
QY 637 isSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnV 657  
Db 3802538 GCGCGCGCGGTGGCGCGCGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3802597  
QY 657 alGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheValGlyThrI 677  
Db 3802598 GGGCGCGCGGAGATTCCAGCGGTGGCGCC-----ACCC 3802633  
QY 677 leGlyGlyValThrGlyAlaTyArgSerAspCysGlyThrThrTrpValLeuIleA 697  
Db 3802634 TCACCGGGGCGCACCGGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC- 3802686  
QY 697 snAspAspGlnHisGlnTyrGlyAsnTrpGlyClnAlaIleThrGlyAspHisAlaAsnL 717  
Db 3802687 -----TACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3802729  
QY 717 euArgArgValTrpIleGlyThrAsnGlyArgGly----- 728  
Db 3802730 TCACACCGCGGT 3802789  
QY 729 -----IleValTyrGlyAspIleGlyAlaProSerGly 740  
Db 3802790 GAGCGT 3802830

## RESULT 11

US-09-252-991A-13875/c  
; Sequence 13875, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13875  
; LENGTH: 2319  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13875

Alignment Scores:  
Pred. No.: 3.35e-05 Length: 2319  
Score: 164.00 Matches: 169  
Percent Similarity: 36.56% Conservative: 80  
Best Local Similarity: 24.82% Mismatches: 275  
Query Match: 4.06% Indels: 160  
DB: 40 Gaps: 40  
US-09-917-376-3 (1-740) x US-09-252-991A-13875 (1-2319)  
QY 72 AlaAlaAspProLeuAsnThrAsn-----LysValTrpAlaAlaValGlyMetTyrThr 89  
Db 1837 GCAGCGGTACCGGCAAGCGGCAACCGCGCTGACCGGTGGCGTGCATACCGACCGCGACG 1778  
QY 90 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 109  
Db 1777 GCACGCGGACACCA-----CCGTGG 1757  
QY 110 GlnIleThrProLeuProPheLeuLeuGlyGlyAsnMet-----ProGly 124  
Db 1756 TG-GTCGGCGCC-----GGCGGACGCTTCGAGGTTCCGCTGAACCGCGCG 1713  
QY 125 ArgGlyMetGlyGluArgLeuAlaVal-----AspPro-----AsnAsnAspAsn 139  
Db 1712 CTGACCAATGCGGACGCGTGTGATCGTTACCGACCGCGCGCGGCAACACAGCACC 1653  
QY 140 IleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp-----Ser 157  
Db 1652 CCGGTGACCGTGGAGGCGCGGACACACAGCCCCGCGCGCGCGCGCGCGCGCGCGG 1593  
QY 158 GlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyTyrTrpIleAlaAsnPro 177  
Db 1592 GCGCGGACGCGACGCGTGTGATCGTGTGATCGTGTGATCGTGTGATCGTGTGATCG 1533  
QY 178 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla-----Phe 195  
Db 1532 ACCGATGCGGACGCGC---CAGCGCGACACACCGCGTGTGTGTGTGTGTGTGTGTGT 1476  
QY 196 AspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp 215  
Db 1475 GAGTCTCCGT 1416  
QY 216 Pro-----AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231  
Db 1415 CCGCGCGGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1380  
QY 232 Ala-----ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspPro 249  
Db 1379 GCTCCCGGACTTCCCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320  
QY 250 ValAsnHis-----ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267  
Db 1319 ACGCGGAAAGCGGCGGT 1260  
QY 268 SerSerGlyAspVal-----TrpLysPheSer-----ValThrSerGly 280  
Db 1259 ACCAGCGCGGATGCAACGCGCAACTGGAGCTTACCCCGCGGTAGCAACTGCGGATGCG 1200  
QY 281 ThrTrpThrArgIleSerProValProSerThrThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 1199 ACC-----GTGGTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1149  
QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal-----Ala 316  
Db 1148 ACCTCCATACCGT 1089  
QY 317 ThrGlnIleSerTrpTrp-----ProAspThrIleIlePheArgSerThrAspGlyGly 334  
Db 1088 AGCGAATCAGCGGACTGT 1032  
QY 335 -----AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsn 348



Db 2730 GTAACTTTGAGTGGGCAACTCTGATGGTTATGTCTGTTGAATACGAGATTTTACGTGAT 2789  
 Qy 103 SerAspGlnGlyAla 107  
 Db 2790 GAGGATGTTGCTTCAACTATTCTGTACACATTTCGGATGAAGACCTTAAATCCAGAT 2849  
 Qy 108 ---ThrTrpGlnIleThrProLeuProPheLeuGlyGly 124  
 Db 2850 ACAACCTACACTTATTCTGTCTGTGTGGTGTGGAGAAAGCGGCGAGAAATCCGCCCAAGT 2909  
 Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuThrPheGly 144  
 Db 2910 GAAGCGTTA---AAAGTGACTACATTAGAGAAATGATGAACCTAAGGAACCGGTGAG 2966  
 Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 162  
 Db 2967 GCGCCGAAATTTACGTATAGCTGATATACAGATACACAGTTCACATCACTGGAAT 3026  
 Qy 163 GlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182  
 Db 3027 GCATCTAATGGTTAC 3050  
 Qy 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202  
 Db 3051 TATGAGGTTCTCGTGATGGTGTGGTT 3080  
 Qy 203 GlyGlnAlaSerLysThrIlePheValGlyVal 213  
 Db 3081 GCGGAACACACGGACACATTCATAGACTACTGGATTAGATGCTGTAGGACCTATACG 3140  
 Qy 214 ---AlaAspPro 216  
 Db 3141 TATACGATTGTTCTCTCGGAGATGGCGGCAAAAGTCTGATCCGAGCGAAGCGTTAGAA 3200  
 Qy 216 216  
 Db 3201 GTAACAACCTCAAGAAAAACAGAGGAATCTAGTAACAATATATATAAAAAAGGCTTT 3260  
 Qy 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGly 235  
 Db 3261 GATACCCCATATATGATATATCTCGGAAGTGGAGAGTGGAGATGCTTCCAGGAAT 3320  
 Qy 235 235  
 Db 3321 AGAATGAAGAAATCAGAAATAGCAGGCTATAGTAAGTTAAACCTGTGATATTCGGAAGCA 3380  
 Qy 236 ---AlaProThrGlyPheIleProHisLysGlyValPhe 251  
 Db 3381 AGCAAGTTGGAAGTAGCCTTTAATAATAGGACGTTGGGCTTGGGATAGTGAAGAGAAT 3440  
 Qy 252 HisValLeuTyr 255  
 Db 3441 AATATTATTTCAGCCAGGTGTTTCATAGTACATTCGCGATCATGAAGAAAGAGGAG 3500  
 Qy 256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrp 273  
 Db 3501 ATTATTCAGGTAAACAGGAGCAGCAATCGATGGTAATAAAGTACGATTACTATCAA 3560  
 Qy 274 ---LysPheSerValThrSerGlyThrTrpThrArgIle 285  
 Db 3561 AATGGCTTTGATACGCCGTATGTTTCATTACCGCCAGAGGCGGAAATTTGGACCAAGCC 3620  
 Qy 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
 Db 3621 CCAGGATTAATAATGGAAGAT 3674  
 Qy 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerThrTrpProAspThr 325  
 Db 3675 GATATTGGTGAAGCTAATCGTCGAGAGTGGCT 3707  
 Qy 326 IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspThrSer 345

Db 3708 ---TTCAATAACGAGCGGGC 3743  
 Qy 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeu 363  
 Db 3744 GAAAAATAAT 3791  
 Qy 364 ---ThrPheGlyValGlnProAsnProProValProSer 375  
 Db 3792 AAAACCGTTTCAGCTGGAGAGATTTATGGAGGTAAAGCAACAGCCATTTAGTGAAT 3851  
 Qy 376 ---ProLys 377  
 Db 3852 GAAGTAATCATTTTATTATAAAATGGTTTTCATACACCGTATGTTTCATATCGTCCAGAA 3911  
 Qy 378 LeuGly 384  
 Db 3912 GGTGTCAGTGGCAAAATGCACAGGAATAAAATGGATAAGTCAGAAATAGCAGGTAC 3971  
 Qy 385 ---MetAlaIleAspProPheAsnSerAspArgMetLeuTyr 398  
 Db 3972 AGTAAATAACGCTTGTATTTGGTCGCGCAGATCGAGTAGAAGTAGCCTTTAATGACCGT 4031  
 Qy 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGly 415  
 Db 4032 CGTGGTGA 4061  
 Qy 416 ---GlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429  
 Db 4062 AATATCTCTTTGTAGTCGGTAACTATCTATGAACAGGAATTAACGGCCACCTGGT 4121  
 Qy 430 ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGly 449  
 Db 4122 CAGGTAAACATCGCTGTACCTGATGATGAGAGATCCG 4166  
 Qy 450 AspLeuGlyGlyPheThrHis 463  
 Db 4167 GATATGAAGACCTGTATACCTCCCTTCAAAGCGAGCTGATTTAAGCAATAGCT 4226  
 Qy 464 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483  
 Db 4227 ACTGCTCATACTGTTTCTAATAGCTGGACA 4280  
 Qy 484 AsnProSerIleValArgAlaGly 492  
 Db 4281 GCTGGGTACAAAATTTATCGAGATGTTGGAATCGGTGTACTGAATCAACAATTTAT 4340  
 Qy 493 ---SerPheAspProSerSerGlnProAsnAspArg 503  
 Db 4341 ACGGATTCAGGTTAAGCGGAGAAACAACTATAGCTATATATGGTACAAAGCTTATGAT 4400  
 Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
 Db 4401 TCTAATTAATTTCTCGCATTAAGTGAATGAACAAATTTGAACCGCGAGAAACGGGT 4460  
 Qy 524 ValThrThrGlyThrValAlaAlaSerAlaAspGlySer 541  
 Db 4461 GTTGATCCAGGAGGATATGCTTTATTCACGAATCCATCGTTTGGTGAAGGTAAACA 4520  
 Qy 542 AlaPro 553  
 Db 4521 ACGCCATCACAATTTGATGGTGTAAATGACGGGAGATGGACAGATGATGTTGTTGCA 4580  
 Qy 554 ValGly 561  
 Db 4581 ATTGGTATGGTGTGACGACCCACGTTGCTCGGGGACAAATTTGTTCTATGATGAACA 4640  
 Qy 562 ---AlaSer 563  
 Db 4641 CCAATGGACCTTACTACCTATGGGAGCATGGGACCATGAGTACTTGTATCTTCTGTTGG 4700  
 Qy 564 GlnGlyVal 576  
 Db 4701 CAATATGTAGATGTAAACAGATATTATTGACCCAGCTAACGCA 4742

QY 577 ValAsnProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSerThrAspGly 596  
Db 4743 -----GGCTCATCAGCTGGTACCACAATAGCCAGATGATATG 4781  
QY 597 GlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 616  
Db 4782 CCACAAACCAATT-----GCAATTGATACCATCCAGAGCAAGGTGCAACA----- 4826  
QY 617 PheHisAlaValProGlyLysGlyGly-----AspLeuTrp----- 628  
Db 4827 -----CATGATATGTGGGGAAAAAATGGTGTGANTCATTGGGAGGACAGATTACCT 4883  
QY 629 -----LeuAlaAsnSerSerGlyLeuTyHisSer----- 638  
Db 4884 GATTACCAACATAATATCGCATCTAATATGTTCCATTGAGGTATATTTCTAGACAGTT 4943  
QY 639 -----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer 652  
Db 4944 GATGGTGTATTTCTGTTGACGATGGAGGAATAAATTATAA-----ACGGGTGAGGAA 4997  
QY 653 SerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyProAlaValPhe 672  
Db 4998 GCAGGAATTACAGTAAGTTTCTAAA-----GGTAAAGGGTAT----- 5036  
QY 673 ValValGlyThrIleGlyGlyValThrGlyAlaTyArgSerAspCysGlyThrThr 692  
Db 5037 -----TCAACATGTGGGGGTGTATAGTCT-----GATGATGCAGTTGATCCT 5081  
QY 693 TrpValLeuIleAsn 697  
Db 5082 AGTAAACTGTGAAC 5096

## RESULT 13

US-07-998-931-4  
Sequence 4, Application US/07998931  
Patent No. 5304723  
GENERAL INFORMATION:  
APPLICANT: Schmid; Candussio; Bck  
TITLE OF INVENTION: Malopentaoose Producing Amylases  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Collard, Roe & Galgano, P.C.  
STREET: 1077 No. 5304723thern Boulevard  
CITY: Roslyn  
STATE: New York  
COUNTRY: USA  
ZIP: 11576  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/998,931  
FILING DATE: 19921229  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GR 40 17 595.2  
FILING DATE: 31 MAY 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Allison C. Collard  
REGISTRATION NUMBER: 22,532  
REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2  
NAME: Thomas M. Galgano  
REGISTRATION NUMBER: 27,638  
REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2  
NAME: Edward R. Freedman  
REGISTRATION NUMBER: 26,048  
REFERENCE/DOCKET NUMBER: SCHMID ET AL-W2  
TELEPHONE: 516-365-9802  
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 5741 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-998-931-4  
Alignment Scores:  
Pred. No.: 0.000168 Length: 5741  
Score: 163.00 Matches: 177  
Percent Similarity: 29.10% Conservatives: 98  
Best Local Similarity: 18.73% Mismatches: 276  
Query Match: 4.04% Indels: 394  
DB: 1 Gaps: 47  
US-09-917-376-3 (1-740) x US-07-998-931-4 (1-5741)  
QY 47 AlaAsnGlyArgTrpIlePro-----LeuLeuAspTrpValGly 59  
Db 2562 TCTTCAGGAAACAAATTCAGGTCCAAATGAACAGGATTTCTTCATTGATCAGATTGGT 2621  
QY 60 TrpAsnAsnTrpGlyTyraAsnGlyValValSerIleAlaAlaAspProIle----- 76  
Db 2622 TGG-----TACGATGCGGTAAAGTGGCTTGATTCAGATCCTTTTGAAGGAA 2669  
QY 77 -----AsnThrAsnLysValTrpAlaAlaValGlyMet 87  
Db 2670 CCTAAGAGCCTGCGCAACACCTAAGACCTAAGTGTGTTAATGTAACCTGAACACTACT 2729  
QY 88 TyrThrAsnSerTrpAspProAsnAspGly-----AlaIleLeuArgSer 102  
Db 2730 GTAACATTTGAGTGGCGCAACATCTGATGTTATGCTGTTGAATACGAGATTTTACGTGAT 2789  
QY 103 SerAspGlnGlyAla----- 107  
Db 2790 GAGGATGTTGTTGCTTCAACTATTCGTACAACTATTCGCGATGAAGACCTTTAATCCAGAT 2849  
QY 108 ---ThrTrpGlnIleThrProLeuProPheLysLeuGlyGly-----AsnMetProGly 124  
Db 2850 ACAACCTACACTTATCTGCTGAGCTGTTGGAGAGCGCGGAGAAATCCGCCCAAGT 2909  
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyPheGly 144  
Db 2910 GAAGCGTTA---AAAGTGACTACATTAGAAGAAATGATGAACCTACAGGAACCGCTGAG 2966  
QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr-----TrpSer 162  
Db 2967 GCGCCAGAAAAATTACGTATAGCTGATATAACAGATACACAGTTACAATCAACTGGAAT 3026  
QY 163 GlnMetThrAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThrThrGly 182  
Db 3027 GCATCTTAATGGTTAC-----GTAAACAGGA 3050  
QY 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202  
Db 3051 TATGAGGTTCTGCGTATGTTGGTT-----ATT 3080  
QY 203 GlyGlnAlaSerLysThrIlePheValGlyVal----- 213  
Db 3081 GCGGAAACAAACAGGACACATTTCATAGATACTGATTAGATGCTGATAGGACCTATACG 3140  
QY 214 -----AlaAspPro----- 216  
Db 3141 TATACGATTTGTTCTCTCGGAGATGGCGGCAAAAGTCTGATCCGAGCGAAGCGTTAGAA 3200  
QY 216 ----- 216  
Db 3201 GTAACAACTCAAGAAAAACAGAGGAATCTAGTAACAATATATACTATAAAAAAGGCTTT 3260  
QY 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGly--- 235  
Db 3261 GATACCCCATATATGCAATTATCGTCCGGAAGGTGGAGAGTGGACGATCGTCCAGGAATT 3320

QY 235 ----- 235  
Db 3321 AGAATGGAAGAAATCAGAAATAGCAGCTATAGTAAGTTAAACGTTGATATTCGGGAAGCA 3380  
QY 236 -----AlaProThrGlyPheIleProHisLysGlyValPhe-----AspProValAsn 251  
Db 3381 AGCAAGTTGGAAGTAGACCTTTAATAATATGACGTGGGCTTGGGATAGTATGATCAAGAGAAT 3440  
QY 252 HisValLeuTyr----- 255  
Db 3441 AATTATTATTGAGCCAGCGTGTTCATACGTACATTCACGATCATCAAGGAAGAGAGAGAG 3500  
QY 256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrp----- 273  
Db 3501 ATTATTCCAGGTAAACCCAGGAGCACCACATCGATGGTAAATAAGTGACGATTTACTATCAA 3560  
QY 274 -----LysPheSerValThrSerGlyThrTrpThrArgIle 285  
Db 3561 AATGCGTTTGATACGCGGTATGTTTCATTACCGCCCAAGAGCGGAAATTTGGACCAACGCC 3620  
QY 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
Db 3621 CCAGGATTAATAATGGAAGAT-----TCAGATTTCGAAGTTATAGTAGGTTAACGCTT 3674  
QY 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr 325  
Db 3675 GATATTGGTGAAGCTAATCGTCGAGAAGTGGCT----- 3707  
QY 326 IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer 345  
Db 3708 -----TTCAATAACAGGACGCGC-----CTTTGGGATAGTGATAAT 3743  
QY 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeu----- 363  
Db 3744 GAAATAAAT-----TATTTCTCAATATTGGCGATAACACTTATATATACCAGGA 3791  
QY 364 -----ThrPheGlyValGlnProAsnProProValProSer----- 375  
Db 3792 AAAAACGGTTCAGTCGAGAGATTATGAGGAGTAAGCCAGCAACCACTTAGTAGGAAT 3851  
QY 376 -----ProLys 377  
Db 3852 GAAGTAATCATTTATTATAAAATGTTTGTATACACCGTATGTTCAITTCATTCCTCCAGAA 3911  
QY 378 LeuGly-----Trp-----MetAspGluAla----- 384  
Db 3912 GGTGGTACGTGGACAAATGCACCAGGAATAAATGGATAAGTCAGAAATAGCAGGTTAC 3971  
QY 385 -----MetAlaIleAspProPheAsnSerAspArgMetLeuTyr-----Gly 398  
Db 3972 AGTAAATAACGCTTGATATGTCGCGCAGATCGAGTAGAAGTAGCCCTTTAATGACGCT 4031  
QY 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGly----- 415  
Db 4032 CGTGGTGCA-----TGGGATAGTATAACGAACGCT 4061  
QY 416 -----GlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429  
Db 4062 AATTATCTCTTTGTAGTCGGTAACAATATCTATGACCAAGGAATTAACCGCCACCTGGT 4121  
QY 430 ThrAlaValAsnAspLeuSerProProSerGlyAlaProLeuIleSerAlaLeuGly 449  
Db 4122 CAGGTGAACATCGCGTGTCTACCTGATGATGGAAGATCCG-----GGA 4166  
QY 450 AspLeuGlyGlyPheThrHis-----AlaAspValThrAlaValPro 463  
Db 4167 GATATTGAAGACCTCATCATACCTCCCTTCAAAGCCGACTGATTTAAACAGCAATAGCT 4226  
QY 464 SerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAlaGluLeu 483  
Db 4227 ACTGCTCATACTGTTTTCATTAAAGCTGGACA-----GCTTCAGCAGCAGCATGAGAATGA 4280

QY 484 AsnProSerIleIleValArgAlaGly----- 492  
Db 4281 GCTGGGTACAAAATTTATCGAGATGTTGGAAATCGTGTACTGAATCAACAACCTTAT 4340  
QY 493 -----SerPheAspProSerSerGlnProAsnAspArg 503  
Db 4341 ACGGATTCAGGCTTAACGGCAGAAACACGTATATAGCTATATGGTACAAAGCTTATGACT 4400  
QY 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
Db 4401 TCTAATAATTTCTCGGCATTAAGTGAATGCAATTTGAAACCCGCGAAGAAACGGT 4460  
QY 524 ValThrThrGlyGlyThrValAlaIleSerAlaAspGlySer-----ArgPheValTrp 541  
Db 4461 GTTGATCCAGGAGGGATATGCTTATTCCACGAATCCATCGTTTGGTAAAGAGTAACA 4520  
QY 542 AlaPro-----GlyAspProGlnProValValValTrpAla 553  
Db 4521 ACGCCAATCACATGATGGTGTAAATGACGGGAATGGACAGATGATATGTTGATGCA 4580  
QY 554 ValGly-----PheGlyAsnSerTrpAla----- 561  
Db 4581 ATTTGGTATGGTGGTACGACCCACGCTTCGTCGGGACAAATTTGCTCTATCATCAAGAA 4640  
QY 562 -----AlaSer 563  
Db 4641 CCAATGGACCTTACTCACCTTATGGGAGCATGGACCATGAGTACTTGTATCTTGTCTGG 4700  
QY 564 GlnGlyVal-----ProAlaAsnAlaGlnIleArgSerAspArg 576  
Db 4701 CAATATGTAGATGAACAGATATATTATGACCCAGCTAACGCA----- 4742  
QY 577 ValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly 596  
Db 4743 -----GCTCATCAGCTGTACCAATAGCCAGCATGATGATG 4781  
QY 597 GlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 616  
Db 4782 CCACAAACCAATT-----GCAATTTGATACCATCCAGAGCAAGGTGCAACA----- 4826  
QY 617 PheHisAlaValProGlyLysGluGly-----AspLeuTrp----- 628  
Db 4827 ---CATGATATGGGGGAAAAATGGTGGTCAATCCTTTGGGAGGAGCCAGATTTACCT 4883  
QY 629 -----LeuAlaAlaSerSerGlyLeuTyrHisSer----- 638  
Db 4884 GATTACCACTAAATATCGCATCTAATATGTTCCATTCAGCTATATTTCTAGACAGCTT 4943  
QY 639 -----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer 652  
Db 4944 GATGGTGTATTTCTCTGTGACGATGGAGGAATAAATTATAAA-----ACGGGTGAGGAA 4997  
QY 653 SerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPhe 672  
Db 4998 GCAGGAATACAGTAAGTTTCTTAA-----GGTAAAGGGTAT----- 5036  
QY 673 ValValGlyThrIleGlyGlyValThrGlyValTyrArgSerAspCysGlyThrThr 692  
Db 5037 -----TCACATATGTTGGGGGTTTAGATGCT-----GATGATGCAAGTTGATCCT 5081  
QY 693 TrpValLeuIleAsn 697  
Db 5082 AGTAAACTTGTGAAC 5096

## RESULT 14

US-09-679-279-1  
; Sequence 1, Application US/09679279  
; Patent No. 6524841  
; GENERAL INFORMATION:  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Volchegursky, Yanina  
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
; TITLE OF INVENTION: Genes and Uses Thereof



```
FILE REFERENCE: 300622004700
CURRENT APPLICATION NUMBER: US/09/679,279
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158,305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190,024
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (928)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDI, deoxysugar transaminase (eryCI, DnrJ homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (5822)...(6595)
OTHER INFORMATION: megDII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megDV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU homolog);
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megAI, KS1
NAME/KEY: misc feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megAI, AT1
NAME/KEY: misc feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megAI, KR1
NAME/KEY: misc feature
LOCATION: (17947)...(18207)
OTHER INFORMATION: megAI, ACP1
NAME/KEY: misc feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megAI, KS2
NAME/KEY: misc feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megAI, AT2
NAME/KEY: misc feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megAI, KR2
NAME/KEY: misc feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megAI, ACP2
NAME/KEY: CDS
LOCATION: (22867)...(33555)
OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (22957)...(24237)
OTHER INFORMATION: megAI, KS3
NAME/KEY: misc feature
LOCATION: (24544)...(25581)
OTHER INFORMATION: megAI, AT3
NAME/KEY: misc feature
LOCATION: (26998)...(27258)
OTHER INFORMATION: megAI, ACP3
NAME/KEY: misc feature
LOCATION: (27393)...(28590)
OTHER INFORMATION: megAI, KS4
NAME/KEY: misc feature
LOCATION: (28897)...(29931)
OTHER INFORMATION: megAI, AT4
NAME/KEY: misc feature
LOCATION: (29953)...(30477)
OTHER INFORMATION: megAI, DH4
NAME/KEY: misc feature
LOCATION: (31396)...(32244)
OTHER INFORMATION: megAI, ER4
NAME/KEY: misc feature
LOCATION: (32257)...(32799)
OTHER INFORMATION: megAI, KR4
NAME/KEY: misc feature
LOCATION: (33052)...(33312)
OTHER INFORMATION: megAI, ACP4
NAME/KEY: CDS
LOCATION: (33666)...(43271)
OTHER INFORMATION: megAII; SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (33780)...(35027)
OTHER INFORMATION: megAII, KS5
NAME/KEY: misc feature
LOCATION: (35385)...(36419)
OTHER INFORMATION: megAII, AT5
NAME/KEY: misc feature
LOCATION: (37068)...(37604)
OTHER INFORMATION: megAII, KR5
NAME/KEY: misc feature
LOCATION: (37860)...(38120)
OTHER INFORMATION: megAII, ACP5
NAME/KEY: misc feature
```

```

; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAIII, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAIII, AT6
; NAME/KEY: misc feature
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAIII, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAIII, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAIII, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megCIII, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
; US-09-679-279-1

```

## Alignment Scores:

```

Pred. No.: 0.00757 Length: 47981
Score: 160.50 Matches: 205
Percent Similarity: 28.61% Conservative: 78
Best Local Similarity: 20.73% Mismatches: 334
Query Match: 3.98% Indels: 374
DB: 4 Gaps: 45

```

US-09-917-376-3 (1-740) x US-09-679-279-1 (1-47981)

Qy	120	yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn-	136
Db	17052	GACCTGTCCGGCCCGC---GAGGACCAAGTCCCTCCGACCCGAGGACGTACGC	17108
Qy	137	-----AsnAspAsn11 140	
Db	17109	CCGCCGTGTGTCAGGGCGGGCGGCGGACGCGCGGTGGAGCCCGCGGGCACCGT	17168
Qy	140	eLeuTyrPheGlyAlaProSerGlyLysGly-	152
Db	17169	GCTCGTCACCGCGGACCGCGGGGTGGTGGCACGTCGCCCGGTGGTGGCCGCCA	17228
Qy	153	-----ArgSerThrAspSerGlyAlaThrTrpSe 162	
Db	17229	GGGCACCCCGTGTGTGGTGGCCAGCGCGGGACCGGACCGCGGGTCGAGCA	17288
Qy	162	rgInMetThrAsnPheProAspValGlyThrTyrIleAla-AsnPro-	180
Db	17289	GCTACTCACCGAACTCCCGACCTGGGACCCCGGGCCACCCTCACCGCTGGCGCTCAC	17348
Qy	180	hrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSers	200
Db	17349	CGACCGGGACGAGCTCC-----GTGCCCTCTCCGCGACCGTCGACGACGAGCA	17396
Qy	200	erSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProv	220
Db	17397	CCCGCTGTCGGCGGTGTTCACGTCGCGGACGCTCGACGACGCGACCGTCGAGACCT	17456
Qy	220	alPheTrpSerArgAspGlyGly-	231
Db	17457	CACCGGTGACCGCATCGAACCGGGCCAAACCGGGCGAAGGTGTCGGTGGCCGCACTGCA	17516
Qy	231	lnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProVala	251
Db	17517	CGAGCTACCGGGACCGCGACCTCG-----AlaThrTrpG 231	17542
Qy	251	snHisValLeuTyrIleAlaThrSerAsnThrGlyProTyrAspGlySerSerGlyA	271
Db	17543	-----ACGCGTTCGTGCTCTCTCTCTCCACCGCGGTCGCGCGCGCGGGG- 17593	
Qy	271	spValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProVal-----	288
Db	17594	-----TCGGCGGTACGTCGCGGGCAACGCTACTCTCGACGG 17630	
Qy	289	-----ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspA	307
Db	17631	TCTCGCCACGACGACGACGAGG-----GACT 17660	
Qy	307	rgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleI	327
Db	17661	CCCGGCCACCTCGGTGGCGGTGACCTGGCGGCGGACGCGGATGGCGGAGGTC--- 17716	
Qy	327	lePheArgSerThrAspGlyGlyAla-	337
Db	17717	-----CGGTGCGCGACCGGTTCGCGCGGACGCGGGTTCATGGAGATGCACCCCGACCGC	17771
Qy	338	--ThrArgIle-----TrpAspTrp-	343
Db	17772	CGTCGAGGGTCTCCGGTGGACCTGGTGGAGGTGAGTAGCCCGATCGTCGTGCGACAT	17831
Qy	344	-----ThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluP	361
Db	17832	CAGGTGGACCGGTTCCTCTCGGTATACCGCGGACGCGCGCCCGCGCTCTTCGACAC	17891
Qy	361	roTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpM	381
Db	17892	CC---TCGACGAGGCGCGTGGGCGCGC---CCGGTCCCGACGCGCGCGGGGTGC 17945	
Qy	381	et-----	381
Db	17946	G-CGCTGGCGGGCTCCCGTCGGGAAACGCGAAGAGCGGTCTCTCGACCTGTGTACGGA	18004
Qy	382	-----AspGluAlaM 385	

Db 18005 CGCAGCGGCTCGCTCGGCCACGCTCGGCCGAGCAGGTGCCGTCGACAGGGCT 18064  
QY 385 etAla-----IleAspProPheAsnSer-----AspArgMetLeuTyrG 398  
Db 18065 TCGCCGAACCTCGCGCTCGACTCGCTGTCGGCCCTGGAACCTGGCAACCGCTGACCACTG 18124  
QY 398 lyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp----- 413  
Db 18125 CGACCGGGTCCGGTGGCCACGACGCGTCTCGACCACCGGAGGTACGACGACCTGG 18184  
QY 414 -----serGlyGlyGlnI 418  
Db 18185 CGGACACCTGCGCCGCAACTGGCGCGGATCGCGCGGAGCGGCCCGCGGCGAG- 18243  
QY 418 leHisIleAlaProMetValLysGlyLeuGluGluThrAlaVal----- 432  
Db 18244 -----GCCCCGAGGTGGCCCCGACGACGAGCCGATCCCATCGTCGGGATGGCCT 18295  
QY 433 -----AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuG 449  
Db 18296 GCCGGTCCCGGGGAGTGGACTCACCGGACAGCTGTGGAGTTGATCGTCCGGGC 18355  
QY 449 lyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrs 469  
Db 18356 GGGACACCGCTCGCGCGGCACCGCGGACCGGAGTGGGATCCGGCGGAGTTGATGGTCT 18415  
QY 469 erProValPheThrGlyThrGlySerValAspTyrAlaGluLeuAsnProSerIleIleV 489  
Db 18416 CCGAC-----ACGACGGGCACCCGTTACCGCTTCGGCAATTCATGCCCGGG----- 18462  
QY 489 alArgAlaGlySerPheAspPro-----Sers 498  
Db 18463 -----GCGGCGAGTTCAGCGCGGCTTCTCGGATCTCGCGCGGTGAGCGGTGGCGA 18517  
QY 498 exGlnProAsnAspArgHis----- 504  
Db 18518 TGGATCCGACGAGCGGACGCGCTGGAGACCACCTGGGAGCGCTGGAGAACCGCGTA 18577  
QY 505 -----ValAlaPheSerThrAspG 511  
Db 18578 TCCGGCCCGAGTCGTTGGCGGTACGGACACCGCGTGTCTTCGTGGGCGATGCCCATCAG 18637  
QY 511 ly-----GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrGlyG 528  
Db 18638 GTTACGCCACCGCGCCCGAAGCCGAGGAGGAGGTGACCGCTACCTGTTGACAGGCA 18697  
QY 528 lyThrValAlaAlaSerAlaAsp-GlySerArgPheValTrpAlaProGlyAspProGly 547  
Db 18698 ACACCGGAGCTCGCTCCGCTCGGTGCGATCGCTGTTGGGTT--GGAGGGCGCGC 18755  
QY 548 GlnProValValTyrAlaVal-----GlyPhe 556  
Db 18756 GATCACTGTGGACACGGCGTGTTCGTCGTCGCTGTGGCGTTCACGTCGGCGCGGTT 18815  
QY 557 -----GlyAsnSerTrpAlaAlaSer----- 563  
Db 18816 GTTGGTTCGGGAGTGTGCTGGCGGTGGCGGTGGGTGTGGTGTGATGCCCGGCTCC 18875  
QY 564 -----GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
Db 18876 GGAGGTGTCAGGAGTTCCTCCGCGAGGCGCGGTGGCTCGGACGG-----CAG 18926  
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
Db 18927 GTGCAAGCCCTTCTCGGACGAGCGCGACGCTTCGTCT-----GGGGAGGGGTGCGCCT 18983  
QY 601 GlnProValAlaAla-----GlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
Db 18984 CGTGCTGTTCAGCGGTGTTCGCTGCGGTGGCGGAGGCGGCTCGGTGTGGGTGTGGT 19043  
QY 619 AlaValProGlyLysGlu-----GlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635

Db 19044 GGTGGTTCGGCGGTGAATCAGATCGGGCGAGTAAATGGTGTGGCGCGCGCTCGGG--- 19100  
QY 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 655  
Db 19101 -----GGTGGCGCAGCAGCGGTGATTCGGCGCGCGTGGGTCTGTGCGGG 19145  
QY 656 AsnValGlyPheGlyLysSerAlaProGlySerSerTyr-----ProAlaValPheVal 673  
Db 19146 TGTGTCCGTGGGATGTGGTGTGGTGGAGGCGCATCGGACGGGACCGCGTTGGGGA 19205  
QY 674 ValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrp 693  
Db 19206 TCCGGTGGATTCGGGCGTGTGGG----- 19232  
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713  
Db 19233 -----GACGTATCGGTGGGTGGGT----- 19259  
QY 714 HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 19260 -----GGTCCGCTGTGTGGTTCGGTGAAGCGAA 19292  
QY 734 IleGlyGlyAlaProSerGly 740  
Db 19293 TGTGGTTCATGTCACGCGCGC 19313  
RESULT 15  
US-09-252-991A-5730  
; Sequence 5730, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5730  
; LENGTH: 4647  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5730  
Alignment Scores:  
Pred. No.: 0.000254 Length: 4647  
Score: 159.50 Matches: 198  
Percent Similarity: 30.51% Conservative: 90  
Best Local Similarity: 20.97% Mismatches: 284  
Query Match: 3.95% Indels: 372  
DB: 4 Gaps: 57  
US-09-917-376-3 (1-740) x US-09-252-991A-5730 (1-4647)  
QY 43 ArgTrpAspAlaAlaAsnGlyArgTrpIle-----ProLeuLeuAsp-----TrpVal 58  
Db 861 CAATGGAGAGTCGCTTGGCACGTGGTCTCTGATCCCTTCGACAGACGAGGTGGTACT 920  
QY 59 GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla-AlaAspProIleAsnTh 78  
Db 921 GGCTGG---CGATGCGCTTTTCATCGAGTTGCG-----GCCACGACGAGTGTCCGCCA 971  
QY 78 AsnLysValTrpAlaAlaValGlyMetTyr-----ThrAsnSerTrpAsp----- 93  
Db 972 GGGACCAACTGGGCATTTGGCCGAGATGTGCCGCGCGGTAGCCAGTCTGTGTCATCGC 1031  
QY 94 -----ProAsnAspG1 97  
Db 1032 CTTGGTCCCGGTGCCGATCTCGCGGTGCGGACACCGTCTGGTGTGGTCTCGGACTCCTC 1091

QY 97 yAlaIleLeuArg----- 101  
Db 1092 GGCAGCCTGAACCTGGCCGATACGCACTATCAGGCGAAGATCAACAGAGCTCGGGAGG 1151  
QY 102 -----SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheIysLeuGI 119  
Db 1152 GCTGGTGTTCACCGATCAGCTACCGACTGGGGATCAACCCG----- 1194  
QY 119 yGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAs 139  
Db 1195 -----GGAATCCGGTCAGCAGAGCAATGAATG 1223  
QY 139 nIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAl 159  
Db 1224 GATC-----TGCGGCTGGGGCGTATTGTGCGGAGCTCCCGCTG 1265  
QY 159 aThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAs 179  
Db 1266 GACCTGGGGCGCC-----GGCAATTACCTGGGCATGCTGCTGG 1304  
QY 179 pThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSer-- 198  
Db 1305 TACGGCATCGCGAGCGGCACTGTGG-----TGTGACGTGTGATCGTCCCT 1355  
QY 199 -----SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAs 217  
Db 1356 GTGTATCGAGAACTCTCGGCACAAACC-----GTGTCACACGCA 1394  
QY 217 nAsnProValPhe----- 221  
Db 1395 GAACCATGTTGTCAGTGTGTGCGCACCGCACTGGCGATCTTGACCTGGCCAGTGGCCG 1454  
QY 222 -----TrpSerArgAspGly-----GlyAlaThrTrpGlnAlaValProGI 235  
Db 1455 TAACTTACGAAATGTCGCGTACGCGCTCTATACCGCGGAACGAGCGCCCGCATGT 1514  
QY 235 yAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu-- 254  
Db 1515 CGCG--ACGGGATTCAATCAGCCTCGCGTCTGTTCAAT-----GGGTCTGACTTGG 1565  
QY 255 -----TyrIleAlaThrSer-----As 260  
Db 1566 CGCGCTGGCGCGGATTAAGAGTGTCTCAACAGCAGCTACGAGCCTGGTATCCGGA 1625  
QY 260 nThrGlyGlyProTyrAsp-----GlySerSer-----GlyAspValTrpLysPh 275  
Db 1626 ACATGCGCGCAACTCGATATCCGTTGGCGGTGATGTGTTGGCGATCAATGGGCGCA 1685  
QY 275 eSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAs 295  
Db 1686 GAAACTGACAAGTAGC-----GATCCCATTCGCCCTTGCCCGCCAGCGCTGAGTCCG 1739  
QY 295 nAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVa 315  
Db 1740 CAACTGGCTCTGGCGCGCAGGC---AGCGCGGATAGGAGAGTGGCCGACG----- 1788  
QY 315 lAlaThrGlnIleSerTrpThrProAspThrIleIlePhe---ArgSerThrAspGlyGI 334  
Db 1789 -----GCTTGGTGGGTCAATTTCGCGACGTATGTGCGTGGCGGAGGAGAGA 1835  
QY 334 yAlaThrTrpThrArgIleTrpAspTrpThr----- 344  
Db 1836 TGTCTCCTAT-----CTGCTGGCTTACCGGTTTCGGCACTCTGGGAGCGGCAATCT 1889  
QY 345 -----SerTyrPr 347  
Db 1890 CAGCATGCGAACCGGAGGTGATCGCGCAATATCGCGCCACGCGGAGATGGCTCGATACC 1949  
QY 347 oAsn-----ArgSerLeuArgTyrValLeuAspIleSerAlaGluPr 361  
Db 1950 TTGCTCGGTAACTTAATCCCGGCGACGAGGACTGGTCTTGGCGGTTCGCGGCACTGG 2009

QY 361 oTrpLeuThr-----Ph 365  
Db 2010 CCGCTCACCTCCACGCGCGCTGCAACTTGGCGGCGGTGACCTGAATGTGAGAAT 2069  
QY 365 eGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMe 385  
Db 2070 TGGCGGGAGGTGAACCCG-----AGCGGGAGGCGCG 2102  
QY 385 tAlaIleAspProPheAsnSerAspArgMet-----LeuTyrGlyThrGly----- 400  
Db 2103 GGCTACCAGACCTACTCGATTGCGGGTTTACCGGTCTTTACAGCGGCGGACCATTCA 2162  
QY 401 -----AlaThrLeuTyrAlaThrAsnAs 408  
Db 2163 CGATCTGAGGGGGCCCTCATCAATCTGCGCGGTTCCGCGAGCTTGATC----- 2211  
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro-----MetVa 424  
Db 2212 -----AGTGTGCGCTGGGAGGATCGATCTCGCTACGACACTCTACT 2255  
QY 424 lLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro-----Pr 439  
Db 2256 GCGGACCCCGCCAGTGGCCAGCAGGAGTGGTTAGCCCCGACGCTGGCTAGCTCCAC 2315  
QY 439 oSerGlyAlaProLeuIleSer-----AlaLeuGlyAsp-- 450  
Db 2316 GGTGGCTCTACCTGCTGGCGGAGATACTGGCATGCGTCTGGAACCCGCTGGCGACCT 2375  
QY 451 -----LeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIle---PheTh 468  
Db 2376 GGTCTGGTGGGTGAGTACCCCGTGGGTGGGTGGGATACCGTATGGCTTTCAC 2435  
QY 468 rSerPro-----ValPheThrTh 474  
Db 2436 GGCTCCGAGTGGCAGCGTCTACCAAGGGGTGGGATCGTGGTTCAGTCTCTGACTGC 2495  
QY 474 rGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPh 494  
Db 2496 ACACACTCGATCGAT-----CTGTTGGCGGCGGCGCAACCT 2534  
QY 494 eAspProSerSerGln----- 499  
Db 2535 GACGCCAGCACCAATTTGGTCGAGGCAACCAACGCTATTCCTATGGCAGGCGCAACCT 2594  
QY 500 -----ProAsnAspArgHisValAlaPheSerThrAsp----- 510  
Db 2595 GTCGCCCTAGCATGGCGATCTATTCATCATAGTAGTAGTGGCGGTGCCCCGGAAGG 2654  
QY 511 -----GlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrTh 526  
Db 2655 AAGTATCTATCTCGTCTCTCCAGCGGTTACATGGGAGCGGTAGCCCTGAATGTCTCGAC 2714  
QY 526 r-----GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 543  
Db 2715 CACTCCCTACTCTTGTGTACTGGCACCGCTCGCTCAACGGTGAACCTGGTCTCTGCT-- 2772  
QY 543 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 563  
Db 2773 -GGAGACTCC-----ATTATGCGCGTGGCTAT-----TCGTCACAGCTC 2813  
QY 563 rGlnGlyValProAlaAsnAlaGlnIle-----ArgSerAs 575  
Db 2814 AGGTGCGCATCTCGGAACCTGCCATCGATCTGGACGCGGCAATTTGCGGCTACTCCGA 2873  
QY 575 pArg-----ValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSe 593  
Db 2874 TGCGCGCTGTCAACCG-----ATAGCGGGCAACCGC-----AG 2909  
QY 593 rThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaVa 613  
Db 2910 TCCGACCGGAAT-----CCGCGCGTCAATTGGCGGATTCCT----- 2946  
QY 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633

```
Db 2947 -----CTGTTCTACTCGGCCCGATAGCGCTGGAGCCCTGGCGCGGATCTGCAACC 2999
QY 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653
Db 3000 GCGCGGTTTCTATCGTTGACAGAGAT-----ATCGTCGGGCTCAATAG 3044
QY 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673
Db 3045 CGGTGCGCAGATTCTGTTTCGGCGAGCAGCG----- 3075
QY 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTr 693
Db 3076 -----GGGAACCGCGCGCGGCAACCTGGTATGAAGTGGGGTCCCGTCTG 3122
QY 693 pValLeuIleAsnAspAspGlnHisGlnTyrGlyAsn---TrpGlyGlnAlaIle----- 710
Db 3123 GATGCGCGCGCGCGACATCGTTGCCCTCCGGAACCCCATTTGGACACGCTATCAGCGC 3182
QY 711 -----ThrGlyAsp-----HisAlaAsnLe 717
Db 3183 TCCAGTCAGATCAGTACGATGCTCGTTACCGCAACCTCTTCGTCCACGACGCC 3242
QY 717 uArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAl 737
Db 3243 CAACGACCTCTCGCTGGTCCAGGCGCGCGCGACATCCTCTACGGCAACTTCAACGTCGC 3302
QY 737 aProSerGly 740
Db 3303 CGGCCCCGGT 3312
```

Search completed: May 12, 2004, 15:15:19  
Job time : 12944.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 1136.23 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTYWNSVAIGGGGFVD.....YIGTNGRGIYVIGGAPSG 740

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QWMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376@cgn\_1\_809@runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1

US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)

1	4036	100.0	2869	13	US-09-917-376-2	Sequence 2, Appli	
2	4036	100.0	2869	15	US-10-153-400-2	Sequence 2, Appli	
3	2429.5	60.2	2646	15	US-10-156-761-1845	Sequence 1845, Ap	
4	2429.5	60.2	9025608	15	US-10-156-761-1	Sequence 1, Appli	
5	1579	39.1	2517	15	US-10-026-994-4	Sequence 4, Appli	
6	1579	39.1	2710	15	US-10-026-994-1	Sequence 1, Appli	
7	1405	34.8	2217	15	US-10-156-761-2561	Sequence 2561, Ap	
C	8	1405	34.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
9	1139	28.2	3668	10	US-09-927-827-21	Sequence 21, Appli	
10	1117	27.7	2367	13	US-10-395-241-13	Sequence 13, Appli	
11	1117	27.7	2646	13	US-10-395-241-11	Sequence 11, Appli	
12	1114	27.6	2481	13	US-10-395-241-17	Sequence 17, Appli	
13	878.5	21.8	5698	13	US-10-420-191-1	Sequence 1, Appli	
14	238	5.9	7407	13	US-10-282-122A-30151	Sequence 30151, A	
15	238	5.9	7407	15	US-10-246-330-3	Sequence 3, Appli	
16	199.5	4.9	7305	13	US-10-282-122A-11269	Sequence 11269, A	
C	17	195	4.8	5295	13	US-10-282-122A-11871	Sequence 11871, A
18	192	4.8	9903	13	US-10-282-122A-28185	Sequence 28185, A	
19	188.5	4.7	4393	16	US-10-369-493-37642	Sequence 37642, A	
20	184	4.6	3324	13	US-10-282-122A-26288	Sequence 26288, A	
C	21	182	4.5	69350	12	US-10-041-018-379	Sequence 379, App
22	180	4.5	2355	16	US-10-369-493-37792	Sequence 37792, A	
23	179.5	4.4	3222	13	US-10-282-122A-14432	Sequence 14432, A	
24	176	4.4	32367	15	US-10-158-160A-14	Sequence 14, Appli	
C	25	170.5	4.2	7191	13	US-10-282-122A-35048	Sequence 35048, A
26	170.5	4.2	25165	15	US-10-114-170-39	Sequence 39, Appli	
27	170	4.2	4323	13	US-10-282-122A-8383	Sequence 39, Appli	
28	169.5	4.2	3360	15	US-10-282-122A-8383	Sequence 8383, Ap	
29	169	4.2	69350	12	US-10-041-018-379	Sequence 11, Appli	
C	30	168	4.2	5451	15	US-10-101-510-597	Sequence 379, App
C	31	167.5	4.2	2110	13	US-10-027-632-97984	Sequence 597, App
C	32	167.5	4.2	2110	16	US-10-027-632-97984	Sequence 97984, A
33	167	4.1	1839	16	US-10-369-493-34868	Sequence 97984, A	
34	166	4.1	3879	13	US-10-282-122A-8717	Sequence 34868, A	
35	165.5	4.1	4383	15	US-10-282-122A-8717	Sequence 8717, Ap	
36	165	4.1	6615	13	US-10-245-802-21	Sequence 21, Appli	
37	164.5	4.1	3237	12	US-10-282-122A-28180	Sequence 28180, A	
C	38	163	4.0	5145	9	US-10-152-319A-1871	Sequence 1871, Ap
C	39	163	4.0	5145	10	US-09-925-299-206	Sequence 206, App
C	40	161.5	4.0	4812	15	US-09-925-299-206	Sequence 206, App
C	41	161.5	4.0	5432	14	US-10-084-817-4	Sequence 4, Appli
42	161	4.0	2712	15	US-10-044-090-22	Sequence 22, Appli	
C	43	159.5	4.0	3129	16	US-10-156-761-3375	Sequence 3375, Ap
44	159.5	4.0	10419	14	US-10-369-493-42336	Sequence 42336, A	
45	159.5	4.0	10419	14	US-10-027-806-3	Sequence 3, Appli	
					US-10-034-623-3	Sequence 3, Appli	

OTHER INFORMATION: a, c, t, g, other or unknown  
US-09-917-376-2

## Alignment Scores:

Pred. No.:	0	Length:	2869
Score:	4036.00	Matches:	740
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-09-917-376-3 (1-740) x US-09-917-376-2 (1-2869)

QY	1	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp	20
DB	139	CGCAGACTCAGCCGTACACCTGGAGCAACGTTGGCGATCGGGGGCGGGCTTTGTCCGAC	198
QY	21	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly	40
DB	199	GGGATCGCTCTCAATGAAGGTGCACCGGAATTCGTACGTGCGGACGACATCGGGGGG	258
QY	41	MetTyrArgTTPAspAlaAlaAsnGlyArgTTPilleProLeuLeuAspTTPValGlyTTP	60
DB	259	ATGTATCGATGGATGCGCCCAACGGCGGTGGATCCCTCTTCGTGATTGGTGGGATGG	318
QY	61	AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	80
DB	319	AACAATTTGGGGGTACAACGGCGTCTGACGATTCGGCGAGACCCGATCAATACTAACAAG	378
QY	81	ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu	100
DB	379	GTATGGGGCGCGTCGGAATGTACACCAACAGCTGGGACCCCAACAGCGGACGATTCCTC	438
QY	101	ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGlyGly	120
DB	439	CGCTCGTCTGATCAGGGCGCAACGTGGCAATAACGCCCTTCGCGCTTCAAGCTTGGCGGC	498
QY	121	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	140
DB	499	AACATGCCCGGCGTGAATGGCGAGCGGCTTGGGTGGATCCAAACAATGACAAACATT	558
QY	141	LeuTyrPheGlyAlaProSerSerGlyGlyLeuTTPArgSerThrAspSerGlyAlaThr	160
DB	559	CTGTATTTTCGGCGCCCGGAGCGCAAGGGCTCTGGAGAAGCACAGATTCGGCGCGGACC	618
QY	161	TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	180
DB	619	TGTTCCAGATGACGAATTTTCGGACGTAGCAGTACATTCGAATCCCACTGACACG	678
QY	181	ThrGlyTyrGlnSerAspIleGlyValValTTPValAlaPheAspLysSerSerSer	200
DB	679	ACCGCTATCAGAGCGATATTCAAGCGCTGCTGGGTGCGTTTCGACAAAGTCTTCGTCA	738
QY	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220
DB	739	TCGCTCGGGCAAGCGAGTAAGACCAATTTTGTGGGCGTGGCGGATCCCAATAATCCGGTC	798
QY	221	PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe	240
DB	799	TTCTGGACAGAGCGGGCGCGGACGTTGGAGGCGGTGGCGGGTGGCGCGCGGCTTC	858
QY	241	IleProHisLysGlyIleValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260
DB	859	ATCCGCACAAAGGGCTCTTTGACCCGCTCAACACGCTGCTATATTGCCACACGCAAT	918
QY	261	ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGly	280
DB	919	ACGGGTGGTCCGTATGACGGAGCTCCGCGACGCTCGAAATTCGAGTACCTCCCGG	978
QY	281	ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr	300
DB	979	ACATGGACCGGAATCAGCCCGGTACCTTCGACGGACACGGCCCAACGACTACTTTGGTTAC	1038

QY	301	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	320
DB	1039	AGCGGCTCACTATTCGACCGCCAGCACCCGAAACAGATATATGTGGCAACAGATATCG	1098
QY	321	TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle	340
DB	1099	TGTGGCGGACACCATATCTTTCGGAGACCCGACGGCGGTGCGAGCTGGACGGCGGATC	1158
QY	341	TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu	360
DB	1159	TGGGATTGACGAGTTATCCCAATCGAAGCTTCGGATATGTGCTTGACATTTCCGCGGAG	1218
QY	361	ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP	380
DB	1219	CCTTGGCTGACCTTCGGCGGTACAGCGAATCCTCCCGTACCCAGTCCGAAAGCTCGGCTG	1278
QY	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetIleuTyrGlyThrGly	400
DB	1279	ATGGATGAAGCGATGGCAATCGATCCGTCAACTCTGATCGGATGCTCTACGGAACAGGC	1338
QY	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle	420
DB	1339	GGGACGTTGTACGCAACAATGATCTCAAGAAGTGGGACTCCGGCGGCGAGATTCATATC	1398
QY	421	AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer	440
DB	1399	GGCGCGATGCTCAAGGATTTGGAGGAGACGGCGGTAAACGATCTCATCAGCCCGCTCT	1458
QY	441	GlyAlaProLeuIleSerAlaIleuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
DB	1459	GGCGCGCCCTCATCAGCGCTCTCGGAGACCTCGGCGGCTTCCACCACGCGCGATTTACT	1518
QY	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
DB	1519	GGCGTGCATCAGCAATCTTTCAGCTCACCGGTGTTCCAGCCGCGCACCGCTCGACTAT	1578
QY	481	AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro	500
DB	1579	GGGGAATTCGATCCGTCGATCATCGTTCGCGCTGGAAGTTTCGATCCATCGAGCAACCG	1638
QY	501	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGlu	520
DB	1639	AACGACAGCACGCTCGCGTTCTCGACAGACGGCGGCAAGAACTGGTTCCAAGGCGAGCA	1698
QY	521	ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal	540
DB	1699	CCTGGCGGGGTGACAGCGGCGGACCGTCGCGCATCGGCGACGCGCTCTCGTTTCGTC	1758
QY	541	TTPAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTTP	560
DB	1759	TGGGCTCCCGCGATCCCGGTGAGCTGTTGTTGTCGAGTTCGGAATTTGGCAATCTCTGG	1818
QY	561	AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys	580
DB	1819	GCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCGCTCAGACCGGGTGAATCCAAAG	1878
QY	581	ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe	600
DB	1879	ACTTTCTATGCCCTATCCCAATGGAACCTTCTATCGAAGACGACGCGCGGCGTACATTC	1938
QY	601	GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal	620
DB	1939	CAACCGGTCCGCGCGGTCTTCGAGCAGCGGTGCGCTCGGTGTCATGTTCCACGCGGTG	1998
QY	621	ProGlyLysGlyAspLeuTTPLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn	640
DB	1999	CCTGGAAGAAAGAGGCGATCTGTGGCTCGCTGATCGAGCGGGCTTTACCACTCAACCAAT	2058
QY	641	GlyGlySerSerTTPSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly	660
DB	2059	GGCGGACAGATTTGTTGTCGAATTCACCGGCGTATCTCCGCGGTGAACGTTGGATTGGT	2118
QY	661	LysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyVal	680



```

Db      2119  |||||AAGTCTGCGCGCGGTCTGTATACCCAGCGCTCTTTGTCGCGCACATCGGAGCGGT 2178
Qy      681  ThrGlyAlaTyrArgSerAspCysGlyThrThrTyrValLeuLeuIleAsnAspGln 700
Db      2179  ACGGGGGGTACCGCTCCGACGACTGTGGAGCAGCTGGGTACTGTATCATGATGACCA 2238
Qy      701  HisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db      2239  CACCAATACGGAATTTGGGACAGCAATCACCGGTGACCAACGCAATTTACGGCGGTG 2298
Qy      721  TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGly 740
Db      2299  TACATAGGCACGAACCGCGGTGAATTTATACGGGACATTTGGTGTGCGCGCTCGGA 2358

RESULT 2
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication NO. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2

Alignment Scores:
Pred. No.: 0 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-3 (1-740) x US-10-155-400-2 (1-2869)
Qy      1  AlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 20
Db      139  GCGACGACTCAGCGCTACACCTGAGCAACCTGGAGCACTGGCGATCGGGGCGCGGTCTTGTCCAG 198
Qy      21  GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
Db      199  GGGATCGTCTTCAATGAAGAGGTGCACCGGGAATTCGTACGTGCGGACCGACATCGGGGGG 258
Qy      41  MetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyrValGlyTyr 60
Db      259  ATGTATCATGATGGATGCCGCCNACGGGCGGTGGATCCCTCTTCGTGATTTGGTGGGATGG 318
Qy      61  AsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db      319  AACAAATGGGGGTACAACGGCGTGTGAGCATTTGGGAGACCCCGCATCAATACTAACAA 378
Qy      81  ValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnAspGlyAlaIleLeu 100
Db      379  GTATGGCGCGCGTCCGAATGTACACCAACAGCTGGGACCCCAACACGAGCGGATTCCTC 438
Qy      101  ArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLysLeuGly 120

```

```

Db      439  CGTCTGTGATCAGGGCGCAACGTGGCAATAACGCCCCCTGCGCTTCAAGCTTGGCGGC 498
Qy      121  AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db      499  AACATGCCCGCGCGTGAATGGCGAGCGCTTGGGTGGATCCAAACAATGACAACATT 558
Qy      141  LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
Db      559  CTGTATTTTCGGCGCCCGAGCGCAAGGCTCTGGAGAAGCACAGATTCCCGCGCGACC 618
Qy      161  TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db      619  TGGTCCCAGATACCAACTTTCCGAGCGTAGCAGCTACATTGCCAATCCCACTGACAG 678
Qy      181  ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSer 200
Db      679  ACCGGCTATCAGAGCGATATTCAAGCGCTGCTGGGTGCTTTCGACAAGTCTTCGTCA 738
Qy      201  SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnProVal 220
Db      739  TCGCTCGGCGCAAGCGAGTAAGACCAATTTTGTGGCGTGGCGATCCCAATAATCCGGTC 798
Qy      221  PheTyrSerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAlaProThrGlyPhe 240
Db      799  TTCGAGCAGAGCGCGCGCGCGAGCTGGCAGCGGTCCCGGTGCGCGACCGGCTTC 858
Qy      241  IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db      859  ATCCCGCACAAGGGGTCTTTGACCGGTCAACACGCTGCTCTATATTGCCACAGCAAT 918
Qy      261  ThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSerValThrSerGly 280
Db      919  ACGGTGTGTCGTATGACGGGAGCTCCGCGCAGCTCTGGAATTTCTCGGTGACTCCGGG 978
Qy      281  ThrTyrThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db      979  ACATGCGACGGAATCAGCCCGGTACTTTCGAGCGACACGCGCAACGACTACTTTGGTTAC 1038
Qy      301  SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db      1039  AGCGGCTCTACTATCGACCGCAGCACCCGAAACACGATATGTTGGCAACCCAGATATCG 1098
Qy      321  TrpTyrProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTyrTrpArgIle 340
Db      1099  TGGTGGCGCGACACCATATCTTTCGAGCAGCGCGGTCCGACGTGGAGCGGATC 1158
Qy      341  TrpAspTyrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db      1159  TGGGATTGGACGAGTTATCCCAATCGAGCTTGGCATATGTGCTTGACATTTTCGGCGGAG 1218
Qy      361  ProTyrLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr 380
Db      1219  CTTGCGGTGACTCTCGGCGTACAGCGAATCTCTCCGATCCCAAGCTCGGCTGG 1278
Qy      381  MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db      1279  ATGGATGAAGCGATGGCAATCGATCGTTCACCTCTGATCGGATGCTCTACGGAACAGGC 1338
Qy      401  AlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIle 420
Db      1339  GCGACGTTGTACGCAACAAATGATCTACGAAGTGGGACTCCGCGCGCGCAATTCATATC 1398
Qy      421  AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440
Db      1399  GCGCGCGATGGTCAAGGATTTGGAGGAGCGCGGTAAACGATCTCATCAGCCCGCGTCT 1458
Qy      441  GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
Db      1459  GCGCGCGCGCTCATCAGCGCTCTCGAGACCTTCGGCGGTTCACCCACGCGCGGTACT 1518
Qy      461  AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db      1519  GCCGTGCCATCGACGATCTTCAGTCCACCGGTGTTTCAACGCGGCGCACCGCGTCACTAT 1578

```

```
QY 491 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGAAATGAATCCGTCGATCATCGTCGCGTGAAGTTTCGATCCATCGAGCCAAACCG 1638
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGlu 520
Db 1639 AACGACAGCGAGTCGCGGTTCTCGACAGCGCGCAAGAACTGGTTCCAGGCGAGCGAA 1698
QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db 1699 CCTGCGGGGTGACGACGGCGGCACCGTCGCGCATCGCGCGACGGCTCTCGTTTCGTC 1758
QY 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db 1759 TGGGCTCCCGCGGATCCCGGTGAGCTGTGTGTACGCGAGTCGGATTTGGCAACTCCTGG 1818
QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGCTTCGCAAGGTGTTCCCGCCCAATGCCAGATCCCGCTCAGACCGGGTGAATCCAAAG 1878
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTTCATGCCCTATCCATGGAACTTCTATCGAAGCAGCGGCGGTGACATTC 1938
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620
Db 1939 CAACCGGTGCGCGCGGTCTTCCGAGCAGCGGTGCGGTGATGTTCCATGTTCCACGCGGTG 1998
QY 621 ProGlyGlyGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTyrHisSerThrAsn 640
Db 1999 CCTGGAAGAAGAGCGATCTGTGGCTCGCTGATCGAGCGGGCTTTACCACTCAACCAAT 2058
QY 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db 2059 GCGGCGACAGTTGGTCTGCAATCACCGCGGTATCTCTCGCGGTGAAAGTGGATTTGGT 2118
QY 661 LysSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyVal 680
Db 2119 AAGTCTGCGCGCGGTGTCATACCCAGCGGTCTTTGTCGTCGCGCACGATCGGAGCGGT 2178
QY 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGln 700
Db 2179 ACGGGGGGTACCGCTCCGACACTGTGGGACGACCTCGGTACTGATCAATGATACACAG 2238
QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db 2239 CACCAATACGGAATTTGGGACAGCAATCACCGGTGACACCGCAATTTACGGCGGGTG 2298
QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 2299 TACATAGGCACGACGCGCGTGAATTTGATACGGGACATTTGGTGGTGGCGGTCGCGA 2358
```

## RESULT 3

```
US-10-156-761-1845
; Sequence 1845, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1845
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2646)
US-10-156-761-1845
```

## Alignment Scores:

```
Pred. No.: 2,58e-229 Length: 2646
Score: 2429.50 Matches: 438
Percent Similarity: 73.99% Conservative: 111
Best Local Similarity: 59.03% Mismatches: 174
Query Match: 60.20% Indels: 19
DB: 15 Gaps: 8
```

US-09-917-376-3 (1-740) x US-10-156-761-1845 (1-2646)

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 115 TACAGCTGGAAGAACGCCCGCGTCGACGGCGGGCTTCGTCGCCGCGCATCGTCTTCAAC 174
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
Db 175 CGCTCCGAGAGAAACCTCGCTACGCCCGCACCGACATCGCGCGCGCTACCGCTGGGCC 234
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnTrpGlyTyr 65
Db 235 GAGTCTCTCAAGACCTCGGACCGCGCTGCTCGACTCGGTCTGGAGCGACTGGGGGCGAC 294
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 295 ACGGGTGTCTGAGCTCGCTCCGACTCGGTCCGACCGGACCAAGGTGTACGGCGCGGTC 354
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 355 GGCAGCTACACGAACAGCTGGGACCGCGGCAACCGTGCCGTGCTCAGGTCCGGGCGCGG 414
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 415 GCGCGAGTGTGCAAGAACCGACCTGCGCTTCAAGCTGGCGGGGAACATGCCGGCGCGG 474
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 475 GGCATGGGCGAGCGGCTCGCGGTGCGACCGGACAGGAACAGCGTGTGTATCTCGGCGCG 534
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 535 CCCAGCGGCAAGGGGTGTGGCGGTGCGACCGACTCGGGGGCCCTCCTGGTCCGAGTCCACC 594
QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 595 GACTTCCCAGACCTCGGCACCTACGTGCGAGGACCGGACCGACAGCGGGGTACCGCGTCC 654
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205
Db 655 GACAAACAGGCGATCGTGTGGGTGTCACCTTCGACGAGTCCGAGGGGTGCCGGGAGGTCC 714
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
Db 715 ACGCGGACGGTGTACGTGCGGGTTCGCCGACAGGACAACTCCCGTCTATCGCTCCACCGGAC 774
QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245
Db 775 GCGGGCGCGACCTGTGTCGGGTGGCGCGCCAGCCCGGCGCATCTCCGCCACAGAGGC 834
QY 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265
Db 835 GTGCTGGAGCGCGGAACGGCTGTGTATCTCGCTCGGTACAGCAGGCGGCGGACCGGTAC 894
QY 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 285
```

Db 895 GACGGCGCAAGGACACCTGTGGCGGTACACGAGACGGACCTGTGACCAATC 954  
Qy 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
Db 955 AGCCGGTTCGGGAGCGGACAC-----TACTACGAGTTCAGCGGCTGACCGTG 1005  
Qy 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTrpProAspThr 325  
Db 1006 GACCGGACGATCCGGGACCGGTGATGGAGCTGCTGATACAGTCTCTGGTGGCGGACAG 1065  
Qy 326 IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleThrAspTrpThrSer 345  
Db 1066 CAGCTTCTCCGCTCCACGAGCGCGGCGACCTGGACGAAGGCTGGGACTACACCTCG 1125  
Qy 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPhe 365  
Db 1126 TATCCGAGCCGCTGAACCGCTTACCATGATGTCCTGCTCGCCCTGCTACCTCGG 1185  
Qy 366 GlyValGlnProAsnProValProSerProIysLeuGlyTyrMetAspGluAlaMet 385  
Db 1186 GGAGCGAACCCCGCACCGCGGAGCAGACCCCGAACTCGGTCGATGACCGAGTCCCTG 1245  
Qy 386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 1246 GAGATCGACCGCTTCGATCCGCGCGCATGATGTACGGAACGGCGCGCATCTACCGC 1305  
Qy 406 ThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
Db 1306 ACGGACACCTGACGAACTGGGACAGCGGAGACCGCAAGTTCACCATCAAGCCGATGCGCGG 1365  
Qy 426 GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444  
Db 1366 GGCCTGGAGGAGACCGCGCTCAACGACCTCGCTCGCTCCCTCGCGGCGGCCAGCTG 1425  
Qy 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
Db 1426 TTCAGCGGCTCGGTGACATCGCGGCTTCGGGCACACGGACCTCACCGGTGCGCTG 1485  
Qy 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
Db 1486 CTGATGTACAGTCGCGGCACTTACACGAGACACCGACCTCGATACGCGGAGACCGAC 1545  
Qy 485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504  
Db 1546 CCGGCGACGGTGTGGGTGGCAATCTCGAC-----TCGGTCCG-----CAT 1590  
Qy 505 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyVal 524  
Db 1591 GTGGCGTTCGACGACACCGCGCAACTGGTTCCGGGGCGGACCTTCGGGGGTC 1650  
Qy 525 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 544  
Db 1651 AGCGGGGTGGGACGCTCGCGCGCGCTCCGACGCGAGTCGCTCGTGGAGCCGCGG 1710  
Qy 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
Db 1711 GGCACCGG-----GTGCAGTACACGACCGGGTTCGGCACCTCGTGTGGCGGTCGCGG 1764  
Qy 565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584  
Db 1765 GGCCTCCCGCGCGGCGCATGCTCAGTCCGACCGGCTCGACCCGAGACCTTCTACCGC 1824  
Qy 585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604  
Db 1825 TTCAGGTCCGCGAGGTTCTACGTCAGTTCGAGCGCGGCGGCGACCTTCACGCGGTCCGCG 1884  
Qy 605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623  
Db 1885 GCCACGCGGCTTCCGAGCGGCGACAGC-----GTGCGCTTCAAGCGCGCTGCCCGGACG 1938  
Qy 624 GluGlyAspLeuThrLeuAlaAlaSerSer-----GlyLeuTyrHisSer 638  
Db 638

Db 1939 AAGCGGACATCTGCTGGCGGGCGGCGAGCGGCGGTACGCGGCTGTGGCACTCG 1998  
Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 1999 ACGGACGCGGCGCGCTTCAACCAAGTTCGCCACCGCTCGACCGGCGGACCATCGGC 2058  
Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
Db 2059 TTCGCAAGGGCGGCGGCGGCGCTCGTACCAAGCGCTCTACACCGCGGAGATCGGC 2118  
Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTyrValLeuIleAsnAsp 698  
Db 2119 GGTGTCCCGGATCTTCGGTCGACCGACAGGGCGGAGTGGACCGCGTCAACGAC 2178  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2179 GATGCCACCATGTTGGACGGCGGCGGATCACCGGTGACCCAGGGTCTACGGG 2238  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 2239 CCGGTGTATGTGTGACGAAACGGCGGCGGATCGTCTACGGCGACACCGCGGCTCTTCG 2298  
Qy 739 SerGly 740  
Db 2299 GACGSC 2304  
RESULT 4  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1  
Alignment Scores:  
Pred. No.: 1.16e-224 Length: 9025608  
Score: 2429.50 Matches: 438  
Percent Similarity: 73.99% Conservative: 111  
Best Local Similarity: 59.03% Mismatches: 174  
Query Match: 60.20% Indels: 19  
DB: 15 Gaps: 8  
US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)  
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 2277508 TACAGCTGAAGACCGCGCGGTTCGACGGCGGCGGTTCGTCCCGCATCGTCTTCAAC 2277567  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 45

Db	2277568	CGCTCCGAGAGAACTCGCTACGCCGACCGACATCGCGCGCTTACCGCTGGGC	2277627	Qy	406	ThrAsnAspLeuThrIysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValuys	425
Qy	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65	Db	2278699	ACGGACAACCTGACGAACTGGGACAGCGGAAGCAGTTCCACCATCAAGCCGATCGCGCGG	2278758
Db	2277628	GAGTCTCTGAAGACCTGGACCGCGCTGCTCGACTCGGTGGAGCGACTGGGGGCAC	2277687	Qy	426	GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu	444
Qy	66	AsnGlyValValSerIleAlaAspProIleAsnThrAsnIysValTrpAlaAlaVal	85	Db	2278759	GGCTTCGAGGAGAGCGCGCTCAACGACCTCGCTCGCTCCCTCCGCGCGCCACAGCTG	2278818
Db	2277688	ACGGGTGTCTGAGCGCTCGCTCCGACTCGCTCGACCCGAAACAGAGTGTACGCGCGCGTC	2277747	Qy	445	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Qy	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105	Db	2278819	TTCAGCGCGCTCGGTGACATCGCGGGCTTCGGGCACACGACCTCACACAGGTGCCGTG	2278878
Db	2277748	GGCACGTACACGACAGCTGGGACCCGGGACAGCGTCCGTCTAGTTCGGCGACCGG	2277807	Qy	465	ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn	484
Qy	106	GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg	125	Db	2278879	CTGATGTACACGCTCGCGAACTTCCACGAGCACCAGCCTCGACTACGCGGACCGAC	2278938
Db	2277808	GGCGCGAGCTGGCAGAAGACCGACCTGCGCTTCAAGCTGGCGGGAAACATGCGCGCGCG	2277867	Qy	485	ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Qy	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145	Db	2278939	CGGGGCACCGTGTGGGGTTCGGCAATCTCGAC-----TCGGGTCCG-----CAT	2278983
Db	2277868	GGCATGGGCGAGCGGCTCGCGGTTCGACCCGAAACAGAAACAGCGTGTCTATCTCGCGCG	2277927	Qy	505	ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal	524
Qy	146	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyValaThrTrpSerGlnMetThr	165	Db	2278984	GTGGCGTTCTCGACGGAACAACGGCGCAACTGGTTCCGGGGGGGGGACCCCTTCGGGGGTC	2279043
Db	2277928	CCACGGGCAAGGGGCTGTGGCGTTCGACGACTCGGGGCTCTCTGTTCGACAGTAC	2277987	Qy	525	ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly	544
Qy	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	185	Db	2279044	AGCGGGGTGGGACGCTCGCGCGCGCTCGACGCGCAGTTCGTTCTGTGTGAGCCCGCG	2279103
Db	2277988	GACTTCCCGAAACGTTCGGCACCTACGTGAGGACGCGACACGACGAGCGGTACGCGTCC	2278047	Qy	545	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	564
Qy	186	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla	205	Db	2279104	GGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCTCGTGTGCGGTCCGCG	2279157
Db	2278048	GACAAACCGAGGATCGTGTGGGTTCACCTTCGACGAGTCGACCGGGTTCGGCGGAGCTCG	2278107	Qy	565	GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Qy	206	SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp	225	Db	2279158	GGCTCTCCCGCGCGGGCGATCGTCGAGTCCGACCGGTTCACCGCCGAGACCTTCTACGGC	2279217
Db	2278108	ACCGGACGCTGTACTCGGGTTCGCGCAAGGACACTCCGCTCTATCGTCCACGGAC	2278167	Qy	585	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla	604
Qy	226	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly	245	Db	2279218	TTCAAGTCCGGCAGGTTCACGTCAGTTCGGACGCGCGGGCGACCTTCACGGGTCCGCG	2279277
Db	2278168	CGGGCGCGACCTGGTCCCGGTGGCGCGCCAGCCCGCCATCTCCGCCACAGGGC	2278227	Qy	605	Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys	623
Qy	246	ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr	265	Db	2279278	GCCACGGGCTCGCGAGCGCGCACAGC-----GTGCGCTTCAAGCGCTGCCCGGACG	2279331
Db	2278228	GTGCTGGACCGCGCAACGGCTGTCTGTACTCGCGTACAGCGCAAGCGCGGACCGTAC	2278287	Qy	624	GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Qy	266	AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle	285	Db	2279332	AAGGGCGACATCTGGCTGGCGCGCGCGGACGCGCGCTACCGGGCTGTGGCACTCG	2279391
Db	2278288	GACGGCGCAAGGACAGCTGTGGCGGTACACGACGAGACCGGGACCTGGACGACATC	2278347	Qy	639	ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Qy	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	305	Db	2279392	ACGGACGGCGCGCGCGCTTCACCAAGCTCCGACCGTCCGACCGCGCGACCATCGGC	2279451
Db	2278348	AGCCCGGTTCGGAGGCGCACACC-----TACTACGGCTTCAGCGGGCTGACCGTG	2278398	Qy	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly	678
Qy	306	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	325	Db	2279452	TTCCGCAAGCGCGCGACCGCGCTCTACACAGCTCTACACACGCGCGGAAGATCGGC	2279511
Db	2278399	GACCGGACATCCCGGGACGCTGTGGCGTACAGCTCTCTGTGTGGCGGACACG	2278458	Qy	679	GlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpTrpValLeuIleAsnAsp	698
Qy	326	IleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	345	Db	2279512	GGTGTGCGGGCATCTTCGGGTTCGACCGCAAGGGCGGAGCTGGACCCCGCTCAACGAC	2279571
Db	2278459	CAGCTCTTCGCTCCACGACAGCGCGCATCTCGAGAAAGCGCTGGGACTACACCTCG	2278518	Qy	699	AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Qy	346	TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe	365	Db	2279572	GATGCCACCATGTTGGAGCGGGCGGCGATCACCGGTGACCCCGAGGTCTACGG	2279631
Db	2278519	TATCCGAGCGCTCGAACCGCTTCACCATGGATGCTCTGTCTCGCCCTGGCTACCTGG	2278578	Qy	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro	738
Qy	366	GlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMet	385	Db	2279632	CGCGTGTATGTGTCGACGAACGGCGCGGATCGTCTACGGCGACACCGCGGGCTCTCG	2279691
Db	2278579	GGAGCGAACCCCGCACCGCGGACAGCACCCCGAAACTCGGCTGGATGACCGATCCCTG	2278638	Qy	739	SerGly	740
Qy	386	AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405	Db	2279692	GACGGC	2279697
Db	2278639	GAGATCGACCCCTTCGACTCCGCGGCGATGATACGGAACCGGGCGGACGCTCTACGGC	2278698				

## RESULT 5

US-10-026-994-4  
 ; Sequence 4, Application US/10026994  
 ; Publication NO. US20030113732A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Goedegebuur, Frits  
 ; APPLICANT: Ward, Michael  
 ; TITLE OF INVENTION: ECVI Endoglucanase and Nucleic Acids  
 ; TITLE OF INVENTION: Encoding the Same  
 ; FILE REFERENCE: GC598  
 ; CURRENT APPLICATION NUMBER: US/10/026,994  
 ; CURRENT FILING DATE: 2002-04-30  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 2517  
 ; TYPE: DNA  
 ; ORGANISM: Trichoderma reesei  
 US-10-026-994-4

Alignment Scores:  
 Pred. No.: 1,78e-145 Length: 2517  
 Score: 1579.00 Matches: 325  
 Percent Similarity: 59.26% Conservative: 123  
 Best Local Similarity: 42.99% Mismatches: 264  
 Query Match: 39.12% Indels: 44  
 DB: 15 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-4 (1-2517)

QY 6 TyrTrpSerAsnValAlaile---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 DB 61 TTTTCGGAAGACGTCAGCTCGCGCGCGCGCGCTTCGTCGCCGCGCATCATCTC 120  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
 DB 121 CATCCCAAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGCTGTACCGCCTC 180  
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
 DB 181 AAC--GCCGACGACTCATGACCGCGCTCAGGATGGGATTGTGTATTAATGCCGCTGG 237  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 DB 238 CACAACCTGG-----GGCATCGACGCTGTGCGCTGTGATCCGAGCATCAAAAG 288  
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 DB 289 GTGTATGCCGAGTCGGCATGTATACGAACAGCTGGGATCCGATTAATGGAGCCATCAT 348  
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 DB 349 CGCTCGTCAGACCGCGCGCAACGTGGTCTTCCACCACTTGGCTTCAAAAGTCGGGGGT 408  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 DB 409 AACATGCCAGACCGGAGCGGAGAGCGTCTGGCTGTGATCCGCGCAATCCCAACATC 468  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160  
 DB 469 ATCTACTTTGGTGTCTCGTCTAGGAACCGGCTCTGGAAAGTCTAGGACGGCGGTGACC 528  
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyIleThrTyrIleAlaAsnProThrAspThr 180  
 DB 529 TTTTCCAAAGTCTCGTCTTCCAGCGCAACTGGGAGTACATCCAGACCCGAGTGATTCC 588  
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200  
 DB 589 AACGGGTACAAACAGCGCAAGAGGACTCATGTGGGTAGCTTTCGATCAACACGAGCAGC 648  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219

DB 649 ACACCGGGGAGGACGCTCTCGTATCTTTGTTGGCAGCGCTGATAACATCACTGCTTCA 708  
 QY 220 ValPheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
 DB 709 GTCTATGTGACACGAATGCGGCTCCAGCTGGAGTGTGTGTCGGCGGAGCCAGGAAA 768  
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
 DB 769 TACTTTCTCACAAGCGGAAACTGCGAGCGAGAGAGCGCTTGTATCTGACTTATTC 828  
 QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
 DB 829 GATGSCACAGGCGCGTATGATGGCACACTTGGCTCAGTGTGGAGTACGACATGTGAGG 888  
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 889 GGAACCTTGAAGAAGCATCACCCCTGTCTGTGATCAGATCTA-----TACTTTGGC 939  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 940 TTTGGCGGCTTGGCTCGATTGCAAAAGCCAGGAACCCCTTGTGTTGCTTTTGAAC 999  
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
 DB 1000 TCTTGGTGGCCAGATGCTCAGCTGTTCGGTCGCGCGACTCTGGGACAAACATGGAGCCG 1059  
 QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
 DB 1060 ATCTGGCGGTGGCGAGCTATCCGACTGAGACTTATTACTAGCAGATCTCAACTCCCAAA 1119  
 QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375  
 DB 1120 GCACCGTGGATCAAGAACAACTTTATCGATGTGACGCGGAGTCACCGTCCGATGTC 1179  
 QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
 DB 1180 ATCAAGCGCTCGGCTGGATGATTCAGTCTCTCGAGATTGACCAACCCAGCAGCAACC 1239  
 QY 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414  
 DB 1240 TGGCTCTACGCGCACCGGAATGACATCTTTGGCGGCCAGATCTCACCACATGGGACAG 1299  
 QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
 DB 1300 GCCACATGTCTCAATCAATCACTGGCAGCGCATCGAGGAATCTCCGTCAGGAC 1359  
 QY 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
 DB 1360 CTGGCTCTGCACCGCGGAGCGAGCTATTGGCGCGCATCGGAGACGACACACGCTTC 1419  
 QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471  
 DB 1420 ACCTTTGCGCAGAAACGACCTCGGAGCATCGCGCAGACGCTCTGGGACGCGCCACA 1479  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 DB 1480 TGGGCCACTCGACAGGCTCGACTACGCGGGAACCTCGTCAAGAGCGTGTGTCGCGTC 1539  
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 DB 1540 GGCAACACCGCGGACCAACAG-----GTGGCCATCTCGTCCGAGCGC 1584  
 QY 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla 531  
 DB 1585 GCGCGCAGTCGAGCATCGACTACCGCGCGCACAGCTCCATGAACCGCGCGACGCTGGCC 1644  
 QY 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551  
 DB 1645 TATTGGCGCGAGCGGACGACGATCTCTGTGTGACGCGCTCGTCCGCGC-----GTG 1695  
 QY 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571

```

Db 1696 CAGCGCTCGAGTTCAGGGCAGCTTTGCTCTCGAGCTCCCGCGGGCGCCGTC 1755
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCGCTCGGACAGAACAGCAACAGCGCTCTTACGCGGCTCGGATCGACCTTTTAC 1815
Qy 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTCAGCAAGGACACCGGACAGCGCTTC-----ACGCGGGG---CCCAAGCTGGCG 1863
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGAGGAGCATCCGGATATCCTCTCACCAGCACCGCGGGCAGCTTGATGTC 1923
Qy 630 AlaLaserSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1924 TCGACCGAGCTCGGCATATTCGCTCCACAGACTCGGCGACGACCTTTGGCCAGTCTCC 1983
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCGCTCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGTCTCA--GGCTCGAACTGG 2040
Qy 669 ProAlaValPheValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGCACC-----GCCCGCTCAGGCGCTCGCCTCTACGCCAGT 2091
Qy 687 AspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCTCTCTGGAGGACATCCAGGGCTCCAGGGCTCCGGCTCCATCCAC 2151
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGGTCCCGCGGACGCGGACGCGCGGCAAGTCTACGTGGGCAACCAACGCG 2211
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly 740
Db 2212 CGGGCGGTCTTTACGCTCAGGGAACCGTCGCGCGCGCGACGCGGGCGG 2259

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 1,976-145 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.26% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
DB: 15 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-1 (1-2710)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGGAAGACGTCAAGCTCGGCGCGCGGCGGCTTCGTCGCCGGCATCATCTTC 160

```

```

Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
Db 161 CATCCCAAGACAAAAGGGCTAGCATATGCACGAACAGATATTGGCGGGCTACCGGCTC 220
Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
Db 221 AAC---GCCGAGACTCATGGACCGCGCTACCGGATGGGATGCTGATATGCGCGCTGG 277
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 278 CACAACTGG-----GGCATCGACGCTGTTCGCTTGATCCGCGAGACGATCAAAAG 328
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 329 GTGTATCGCGCAGTCGGCATGTATACGAACAGCTGGATCCGAGTAATGGAGCCATCAT 388
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 120
Db 389 CGCTCGTCAGACCGCGCGCAACGTTGCTCTTCAACCACTTGCCTTCAAAAGTCGGGGT 448
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 449 AACATGCCAGGACGCGGAGCGCGAGAGCTCTGGCTGTGATCCGCCCACTCCACATC 508
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
Db 509 ATCTACTTTGGTGTCTCGCTCAGAAAACGGCTCTGGAAGTCTACGACGCGCGCTGACC 568
Qy 161 TrpSerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 569 TTTTCAAGGTCTCGCTGTTCACGCGCACTGGGACGTATACCTCCACGACCGAGTATCC 628
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
Db 629 AACGGCTAACAGCAGCAACGCAAGACTCATGTGGGTACGTTTCGACTCAACCCAGCAGC 688
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219
Db 689 AGCAGCGGGGAGCCACGCTCTGATCTTTTGGCAGCGCTGATAACATCACTGCTTCA 748
Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
Db 749 GTCTATGTGAGCACCAATGCGGCTCCAGTGGAGTGTGTACCGGGGCGAGCCAGGAAA 808
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValIleTyrIleAlaThrSer 259
Db 809 TACTTTCTCACAGGCGGAAACTGCGAGCCAGCAGAGAGCGCTTGTATCTGACTATTTC 868
Qy 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
Db 869 GATGGCACAGGCGCTATGATGGCACACTTGGCTCAGTGTGGAGGTACACATTCGAGGG 928
Qy 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 929 GGAACCTGGAAAAGACATCACCCCTGTCTCTGGATCAGATCTA-----TACTTTGGC 979
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 980 TTTGGCGGCTTGGCTTCGATTGCAAAAGCCAGGAACCCCTTGTCTTCTTCTTTCAAC 1039
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpTrpArg 339
Db 1040 TCTTGTGGCCAGATGCTCAGCTGTTTGGTGCACCGACTCTGGGACACATGGAGCCCG 1099
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
Db 1100 ATCTGGCGTGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCACTCCCAA 1159
Qy 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375
Db 1160 GCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCAACCGTCCGATGCTC 1219

```



Qy 376 ---ProLysLeuGlyThrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
Db 1220 ATCAAGCGCTCGGCTGGATGATGAGTCTCGAGATTGACCAACCGACAGCAACAC 1279  
Qy 395 MetLeuTVrGlyThrGlyAlaThrLeuThrAlaThrAsnAspLeuThrLysTrpAspSer 414  
Db 1280 TGGCTCTACGGCACCGGATGACATCTTTGGCGCCACGATCTCACCACCTGGGACAG 1339  
Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
Db 1340 CGCCACATGTCTCAATCAATCACTGGCAGACGGCATCGAGGAATTCCTCGTCAGGAC 1399  
Qy 435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
Db 1400 CTGGCTCTGCACCCGCGGAGCGAGCTATTGGCCGAGTCGGAGACGACCAACGGCTTC 1459  
Qy 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471  
Db 1460 ACCTTTGCCAGCAGAAACGCTCGGACATCGCGCAGACGGTCTGGGCAACGCCACCA 1519  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1520 TGGGCCACTCGACGAGCTCGACTACCGCGGAACCTCGGTCAAGAGCGTCTCGCGCTC 1579  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1580 GGCAACACCGCGGACGCAACAG-----GTGGCCATCTCGTCCGACGGC 1624  
Qy 512 GlyLysAsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla 531  
Db 1625 GCGCGAGTGTGAGCATCGACTACGCGGCCGACACGTCATCAACGCGGCGACGGTGCC 1684  
Qy 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551  
Db 1685 TATTCGCGGACGCGGACACGATCTCTGTGCGACCGCTCGTCCGGC-----GTG 1735  
Qy 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571  
Db 1736 CAGCGCTCGAGTTCAGGGGAGCTTGTCTCGCTCGACCTCGCCGCGGCGCGCTC 1795  
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591  
Db 1796 ATCGCTCGGACAGAACAGACGAGCTCTTACGCGCGCTCGGATCGACCTTTTAC 1855  
Qy 592 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611  
Db 1856 GTACAGAGGACACCGGACGAGTTC-----ACGCGCGG---CCCAAGCTGGGC 1903  
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629  
Db 1904 AGCGGAGGACGATCCGGATATCGCTGCTCACCGACACCGCGGGCAGCTTGTATGTC 1963  
Qy 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648  
Db 1964 TCGACCGAGCTCGCATATTCGCTCCACAGACTCGGCGACGACCTTTGGCCAACTCC 2023  
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668  
Db 2024 ACCGCTTCGACCAACACCTACCATCGCTCGGCTGTGGGTCTCA---GGCTCGAACTGG 2080  
Qy 669 ProAlaValPheValValGlyThrIleGlyValThrGlyValAla-----TyrArgSer 686  
Db 2081 ---AACCTGTATGCTTCGGACCC-----GGCCCGTCAGGGGCTCGGCTCTACGCCAGT 2131  
Qy 687 AspAspCysGlyThrThrTrpValIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706  
Db 2132 GGAGACACGGCGCTCTCGACGACATCCAGGCTCCAGGGCTTCGGTCCATCGAC 2191  
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726  
Db 2192 AGCACAAAGTTCGCGGCGGACGACACCGCGGCGAAGTCTACGTGGGACCAACCGGC 2251  
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyAlaProSerGly 740

Db 2252 CGGGCGCTCTTTTACGCTACGGAACCGTCTGGCGGCGGACCGGGCGG 2299  
RESULT 7  
US-10-156-761-2561  
; Sequence 2561, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2561  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2217)  
US-10-156-761-2561  
Alignment Scores:  
Pred. No.: 2,178-128 Length: 2217  
Score: 1405.00 Matches: 308  
Percent Similarity: 54.57% Conservative: 104  
Best Local Similarity: 40.79% Mismatches: 257  
Query Match: 34.81% Indels: 86  
DB: 15 Gaps: 21  
US-09-917-376-3 (1-740) x US-10-156-761-2561 (1-2217)  
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 124 TACCGCTGGCGCAACCGCTCATCGGGGACCGGCTTGTCCACCGCTGCTCTCCAC 183  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
Db 184 CCCTCGTACGCGGTCTCGCTACGCCGCGGACCGACATCGGCGGCGCTACCGCTGGAC 243  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 244 GACCGCGCGCGCGCTGGACCGCCGCTCATCGACCACTCGGCTGGGACGACTGGAACTC 303  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 304 CTCGCGCTCGAGGCGATCGCTGACCCACCGACCGGACCGGCTCTACCTCGCGCTG 363  
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 364 GGCACCTACGCCAGTCTGTGGCGGGCAAC---GGCGGCTCTCGCTCCGAGGACCGC 420  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 421 GCGCGCACCTGACCGCGGACCGACCGACCTGACCGTGAAGCTCGGCGGCAACGAGGACGCGC 480  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 481 GCGCGCGTGTAGAGCTCTCTCTGACCGCGGACGAGCAGCACCTCTGTGCTGGGCAAG 540  
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165



Db 541 ---CGGCACGCGGCTGCTCAAGTCGACCGACCGGGCGCCACTTGGGCGACCGCGGACC 597  
QY 186 AsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
Db 598 GCGTTCCCG-----CGCAAGCGGAACCTCC 621  
QY 186 AspIleGlnValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 622 TCCGGCAGGAGTCTGTGTC-----CTCGTCGCGCGC 654  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
Db 655 GGGCGACCGTCTACCGCGTGGGTGACCGGACCGGCACTCGGGCAGCGGAACCTG 714  
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 715 TACCGCACGGCGGAC---GGCACGACCTGGGGCGGTCCCGCGCGGCGCTCCGGCACCC 771  
QY 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
Db 772 TCCGCCAAGGTCCTCGCGCCGCGTACGACACGACACCGCGGAGCTGTACGTGACG 831  
QY 258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal 277  
Db 832 TACGGGACGACCGCGCGCGCGCGCGTCCGACGCGAGGTGCACAACTGCGTACCC 891  
QY 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
Db 892 GCCCGGAGCGTGGACGAGGTACCCCGGTGAAGCGCGGGGACGACGACGACGCGC 951  
QY 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
Db 952 TCGGGGACACCTTCGCTACGCGCGGGTTCGCGTGCAGCGCGCGCGCGCGCACCTC 1011  
QY 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIlePheArgSerThrAspGly 333  
Db 1012 GTGCTCTCACCAACACCGCTGGGCGGACGCGGACCGGTCTTCGGTCCACGACGCGC 1071  
QY 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 1072 GCGCTTACTGACGTCCTCAAGGAC-----GCCGCC 1104  
QY 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Db 1105 GTGTTCCGAGTCTCCGAGACTCCCTCTCTCGACTGGGCGGACGAC----- 1149  
QY 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaAspProPheAsnSerAsp 393  
Db 1150 ---AAGCCGAAGTTCGGCTGGTGGATCCAGGCGCTCGCGGTGCAGCCCGTACGACCTCC 1206  
QY 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 1207 CACGTCTGTACGGACCGCGCGGACCTCTACGGCACCGGACCTCAAGCGCTGG--- 1263  
QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 1264 -----GCACCGCGATCCGCGGCTGGAGAGCGCGCTGGCGC 1302  
QY 434 AspLeuIleSerProProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
Db 1303 CAACCTGATCTCGCCCGGCTCGGGGAGGACACCTGATCAGCGGACTCGGGACATCGGT 1362  
QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
Db 1363 GTGATGTACCGACGCGCTCAGCGGTCTCGTCCGCGGATGGCGACGACCCCGG 1422  
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1423 TTCGGTCCGCGACGGGACTCGCGAGCGCGCGGCGGTATGCTGCTCGCAGC 1482  
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1483 GGCTGGGCGCACCGGCAACGCG-----GCGTACTCCACGACGCGC 1524

QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
Db 1525 GGGCGACCTGGGCGCGCTTCGAGGCCCGACCGCCGACATCCCAAGGACGACCGCGGACCG 1584  
QY 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
Db 1585 ATCGCCACAGTTCGCGACGCGGCGCACATGCTGTGGTCTCTGTGCACTGGGACGCG--- 1641  
QY 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
Db 1642 -----ACGACGTACGCGGCGCCACCGCTCGACGCGACAAACGCGCGGAGTGTGTC 1689  
QY 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
Db 1690 GAGGTCTCTCTTCCGAAAGGCGCGCACACCGGTGCGCGACCGCGCGATCCGACGCGC 1749  
QY 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599  
Db 1750 TTTTACGCTTACGACTTCGACAAATGGAACGCTATACGCCACGACTGACAGTGGCGCTTCG 1809  
QY 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
Db 1810 TTTACGCGCGGTGCGGCGGACTGCCC---TCCGCGACAGCCAGCTTCAAGCTGCTCGCG 1866  
QY 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638  
Db 1867 GCGCGGGACGAAGCGCGGACCTGTGCTCTCCGCCAAATGGAACGCGGCTTACCGGTCC 1926  
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 1927 ACCGACGGCGGGGACACTTTCGCGAGATCGACAGCTGTGGGCTCTGTACACCTCGGC 1986  
QY 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
Db 1987 TTCGGAAGCGCGCGCGGCGGCTACCGCGGATCTTACGAGTCTGGCTCGGCTCAGCGGAG 2046  
QY 679 GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpTrpValLeuIleAsnAsp 698  
Db 2047 ACCATCACCCTGCTTACCGCTCCGACGCGCGCGGACATGGGTCCGGATCAACGAC 2106  
QY 699 AspGlnHisGlnTrpGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2107 GACGCCACCACTGGGGGTGGATCGCGGAGCGCGTCTCGTGCATCCCGCATCCACGCGC 2166  
QY 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 2167 CGGCTCTACTCGCCACCAACGCGCGCGGATCCAGTACGGGGAG 2211

## RESULT 8

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: misc feature

LOCATION: (4187715)

OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.: 1.24e-123 Length: 9025608  
Score: 1405.00 Matches: 308  
Percent Similarity: 54.57% Conservative: 104  
Best Local Similarity: 40.79% Mismatches: 257  
Query Match: 34.81% Indels: 86  
DB: 15 Gaps: 21

US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25  
Db 3161882 TACCGCTGGCGCAACGCGTTCATCGGGGACCGCGTTCGTCACCGCGTCTTCCAC 3161823  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 3161822 CCCTCGGTACGCGTCTCGCTACGCCCGGACCATCGCGCGGCTACCGCTGGAC 3161763  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 3161762 GACCGCGCGCGCGTGGACCGCGTTCATCGACCATCTCGGCTGGGACGACTGGAACCTC 3161703  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 3161702 CTCGGCGTCGAGCGATGGCGTTCGACCCACGACCGGACCGGCTCTACCTCGCGGTG 3161643  
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 3161642 GGCACCTACGCCCGATGTCGGCGGCGCAAC---GGCGGTCCTCGCTCGGACGCGC 3161586  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 3161585 GCGCCACCTGGACCGCGCACGACTGACCGTGAAGCTCGCGGCAACGAGGACGCGCGC 3161526  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 3161525 GCGCGCGTGAGCGACTCTCTCGACCGCGCGACGACACCTCTGCGTGGCGACG 3161466  
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
Db 3161465 ---CGGCACGACGGGCTCTCAAGTCGACCGACCGGGGCGGCACTTGGCGACGCGGAC 3161409  
Qy 166 AsnProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 185  
Db 3161408 GCCTTCCCG-----GGGAGCGGAACCTC 3161385  
Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 3161384 TCCGGGAGGAGTCGTTC-----CTCGTCCGCC 3161352  
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
Db 3161351 GGGCGCACCGTCTACCGCGGTGGGTGACGCGACGCGCTCGGCGACGCGCAACCTG 3161292  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 3161291 TACCGACGCGCGAC---GGCACACCTGGGGGCGGCTCCCGCGCGCGCTCGGCGAC 3161235  
Qy 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
Db 3161234 TCCGCCAAGTCTCCGCTCCGCGCGGTACGACGCGACACCGCGGAGCTGTAGTGACG 3161175  
Qy 258 ThrSerAsnThrGlyProTyrAspGlySerGlyAspValTrpLysPheSerVal 277  
Db 3161174 TACGCGACGACCGCGCGCGCGCGCGCGTCCGACGCGCGTGCACAAAGCTGGTACC 3161115

Qy 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
Db 3161114 GCCACGGGACGTGGACCGAGGTCAACCGGTGAAGCCGGCGGACGCGGACGACGCGC 3161055  
Qy 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
Db 3161054 TCGCGGACACCTTCGCTACGGCGGGTTCGCGCTGACGCGCGCGCGCGCCGCCCTC 3160995  
Qy 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
Db 3160994 GTCGTCTCCACCAACACCGCTGGCGGACGCGGACGCGTCTTCGCTCCGCGACGCGC 3160935  
Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 3160934 GCGCGTACCTGGACGTCCTCAAGGAC-----GCCGCC 3160902  
Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Db 3160901 GTGTTCGACGTGTCGAGACTCCCTTCCTCGACTGGGCGGACGAC----- 3160857  
Qy 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 3160856 ---AAGCGAAGTTCGGTGTGGATCCAGGCGCTCGCGGTGACCCGTCACGATCCAG 3160800  
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 3160799 CACGTGTTGTCGCGGACGCGCGGACCTCTACGCGACCCCGGACCTCAAGCGCTGG--- 3160743  
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 433  
Db 3160742 -----GCACCGGATTCGCGCGCTGGAGAGAGCGCGTGGC 3160704  
Qy 434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
Db 3160703 CAACTGATCTGCCCGCGTGGGAGGACACCTGATCAGCGGACTCGCGGACATCGGT 3160644  
Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
Db 3160643 GTGATGTACACGACGCGCTCAGCGCTCTCGTCCGCGGACATGCGGACGACCCCGTG 3160584  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 3160583 TTCGGTTCGCGACGCGGACTTCGCGAGCGCGCGGCGCGGCGGTATGTCGTCGCGACG 3160524  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 3160523 GCGTGGGCGACCCCGCAACGCG-----GCGTACTCCACGACGCGC 3160482  
Qy 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
Db 3160481 GGGCGGACCTGGCGCCCTTCGAGGCCCGACCGGACATCGCCCAAGGACGACCGGCGCGC 3160422  
Qy 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
Db 3160421 ATCGCACAGTTCGCGACGCGCGGACACATGCTGTGTGCTCTCTGTCGACTGGGACGCGC 3160365  
Qy 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
Db 3160364 -----ACGACGTACCGCGCCCGCCCGTTCGACGAGCAACCGCGCGAGCTGTCC 3160317  
Qy 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
Db 3160316 GAGGTCTCTCTCCGAGGGCGGCGCACCGGTTCGCGGACCGCGCGCGATCCGACGCGC 3160257  
Qy 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599  
Db 3160256 TTCTACGCTACGACTTCGACCAATGGACGCTATACCGCGGACGACTGACAGTGGCGCTCG 3160197  
Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
Db 3160196 TTCAGCGCGCGTTCGCGCGGACTGCGC-----TCGCGGACGACGCGAGTCAAGTGGTGGC 3160140  
Qy 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638

```
Db 3160139 GCGCCGGGACGAGCGGCGACTGTGGCTCTCCGCCAAATGAAACGGGCTCTACCGGTCC 3160080
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658
Db 3160079 ACCGAGCGGGGACACACTTCGCCAGGATCGACAGTGTGGGCTCTGATACCTCGGC 3160020
QY 659 PheGlyLysSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGly 678
Db 3160019 TTCCGCAAGCGCGCGACGCGCGCGACTACCGCGCGATCTACACAGTCTCGCTCGACGGAG 3159960
QY 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
Db 3159959 ACCATCACCGCGCTTACCGCTCCGACGACGCGCGACGAGCATGGGTCCGGATCAACGAC 3159900
QY 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
Db 3159899 GACGCCACCATGGGGGTGGATCGCGAGGCGTCTCGGTGACCCCGCATCCACGGC 3159840
QY 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 3159839 CGGGTCTACTCCGCCAACCAACGCGCGCGCATCCAGTACGGGAG 3159795
RESULT 9
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication NO. US2003036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21
Alignment Scores:
Pred. No.: 7,17e-102 Length: 3668
Score: 1139.00 Matches: 267
Percent Similarity: 50.85% Conservative: 120
Best Local Similarity: 35.09% Mismatches: 295
Query Match: 28.22% Indels: 79
DB: 10 Gaps: 21
US-09-917-376-3 (1-740) x US-09-927-827-21 (1-3668)
QY 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
Db 1217 GCCAGTCCGGGCGCTTACCAGTGGCGAGTTCGCGATTCGGCGGTGGCGCTTTGTACCC 1276
QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
Db 1277 GGTGTGCTTTTCATCCCGCCGAAAGTGTGTGGCTATCGCGCACCATGATGGGTGGC 1336
QY 41 MetTyrArgTrpAspAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60
Db 1337 GGTACCGGTGGGATGCGCAGCGCAGCAGCGCGCGCTGACCGACTGGTTGGCGCT 1396
QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 1397 GACGACTGAACCTGATGGGATCGACGATTCGCGCTGCGCGTCCAGGACCGTGGGTGGAGAACCGTCCG 1456
QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
```

```
Db 1457 CTGTATCTCGCGCGCGCACTATATGTCATGAA---CGCGCCGGCACTGCGCGAGTGTG 1513
QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
Db 1514 CGTCTGTTCAACCGCGCGCGCAGCTTCGAGCGTCCGACCTGCGCTTTAAGCTGGGTGGT 1573
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 1574 AACCACTGGCGCGCGCAATGCGAGCGGCTGGCGGTGACCGCACAGTATGGCGCGGTG 1633
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyValaThr 160
Db 1634 CTGCTCTGGGCTCGCGCGATGCC---GGCCTGTGGCTAGCGACGATCGCGCGCGCAC 1690
QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 1691 TGGCGCAAGGTGGCGTTCGCGAGCGCGCGCTGGCGGTGCCCGCGCGCAATCAT 1750
QY 181 ThrGlyTyrGlnSerAspIleGlnValValTrpValAlaPheAspLysSerSerSer 200
Db 1751 GTTGGCGCGCAGCAGCGGTG---GGATCGCTTTGTCGTTCGACGCGCAGCGTGGC 1807
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 1808 AACACCGGTTCGCCAACACCGCGCATCTACGTGGCGGTGCCACCGAACACGACGCTG 1867
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 1868 TATGTGTCGAAGATCGCGCGCGCAGTTCGGCACCGGTGGCGGCGGACACCGCGCGCTG 1927
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 1928 CGCCCGAGCCACATGCGCGCGCGCAGCGATGGGCAC---TGGTATCTGAGTATGGCGAC 1984
QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
Db 1985 CAGCCCGCGCGCAGCCTGATGGCGGGGAGCGCTTGTGAAATTCACCGCGCGCACAGGG 2044
QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 2045 CGTGGCGTGGATCAGCCCGATTCGCGAG---CCAGCCAGTGGCGATGATTCGGCTGG 2101
QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db 2102 GGTGGGTGGCGGTGGATCGCAACATCCGACAGTGTGCTGGCGCACACCTTCGCGCGT 2161
QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340
Db 2162 CGCACCGCGCGCAGCAGCTGTATCGCAGCGTGGATGGTGGCAAGCAGTGCAGCGCGTGG 2221
QY 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 2222 -----TTGGCGCATGCGGTGTTCGATCAACGCGCGCG 2254
QY 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380
Db 2255 CCGTGG---ACCGCACATGCCACCGCGCAC-----TGG 2284
QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 2285 ATGGGG---CGCTGGCGATCGATCTCGACGCAACCATCGCTGTTCGTGACCGCGC 2341
QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
Db 2342 TACGCATCTGGGCGCTCGCGCAATCTG-----CAGGATTCGCC 2380
QY 421 AlaPro-----MetValLysGlyLeuGluThrAlaVal 432
Db 2381 GCACCGCAGCGCGCGCTGCGTGGTGGTTCAGGACCGTGGGTGGAGAACCGTCCGCG 2440
QY 433 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452
```

Db 2441 CTGGACCTGCTCAGCCCGATGGTGGCGCGCATCTGCTCAGCGCGCTCGCGGATATCAAC 2500  
Qy 453 GlyPheThrHisAlaAspValThraValProSerThrIlePheThrSerProValPhe 472  
Db 2501 GGCCTTCGGCATGACGACCTGGACCGCGTG--CAGTTGCAGTACGCGCGCCACGGGTG 2557  
Qy 473 ThrThrGlyThrSerValAspTyrAlaGluIleuAsnProSerIleIleValArgAlaGly 492  
Db 2558 ACCAATGGCGAAAGCATCGATGCGCGCGCCAGCGCGCGAGTGGTGGTGGCGAGCGGT 2617  
Qy 493 SerPhe-AspProSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 512  
Db 2618 ACCGTGGCGCGCCCGCCGCAACAGAAATCCGCGCGC---TGTATTACGCTGATGGCGG 2674  
Qy 512 YLysAsnTrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531  
Db 2675 CAAGCAATGGACTCGGTTGGAGTGGAGTGGAGCGCGCGGCGCGCGCGCGCATTTGC 2734  
Qy 531 AlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValVa 551  
Db 2735 CATTGGTCCGATGCGCGCGAGGTGTGTGGCACCGCGCGCGCGCAAT-----TG 2788  
Qy 551 lTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaG 571  
Db 2789 GCGCACGTCGCACTTCGCGCGCGCAGTGGCAGCGTGTGAACGCCCTGCGCGCAACCGCGGTG 2848  
Qy 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589  
Db 2849 GGTGTGACCATTCGGTGGAGCGCACCGCGCGTGGTATCGGTGGATGTCGCCAGCGGCA 2908  
Qy 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu---- 607  
Db 2909 GCTGTACGAGACCGATGCGCGCGCAGTTTCGT-----GCGACCGGTGTACAGGT 2962  
Qy 608 -----ProSerSerGlyAlaVa 613  
Db 2963 GGCAGCGCGCGCGCGATGAGCGCACCGCGCGCGAGCTGCGTCCGCGACCGCGCGCGC 3022  
Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSe 633  
Db 3023 AGCGTGTGTATCTGGCCACCGCGCGCAAG----- 3053  
Qy 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653  
Db 3054 -GGTGTGATGCGTGGCAGGACGT-----GCGTGTGAGTGTCTTCACAGCGCGGACA 3106  
Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673  
Db 3107 AGCAGCTCGCTGGCATCGCAAGCGCTGCGGCGCGTGGCGCGCGCGCGTGTATCT 3166  
Qy 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTr 693  
Db 3167 GCGCGCGCGTGGAGCGGTGGATGGCGTGTCCGCTCCGACGATGGCGCGCGCGCATG 3226  
Qy 693 pValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713  
Db 3227 GCAGCGCATCAACGATGACGCGCACCGCTTCGGCGC---CCGTACAGCGTACCGGTGA 3283  
Qy 713 pHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733  
Db 3284 TCCGCGCATTCGCGCGCGTGTGTACTTCGCCACCGCGCGCGCGCATTTTCTACGGCGA 3343  
Qy 733 p 733  
Db 3344 T 3344

## RESULT 10

US-10-395-241-13  
; Sequence 13, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi

; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 2367  
; TYPE: DNA  
; ORGANISM: Geotrichum sp. M128  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: (1)..(2367)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2367)  
; US-10-395-241-13

## Alignment Scores:

Pred. No.: 5,94e-100 Length: 2367  
Score: 1117.00 Matches: 281  
Percent Similarity: 50.37% Conservative: 124  
Best Local Similarity: 34.95% Mismatches: 279  
Query Match: 27.68% Indels: 120  
DB: 13 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-13 (1-2367)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 10 TACGAGTTCAGATGTCGCGATCGCGCGCGGGGTACATTACCGGGATTGTCGCGCAC 69  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 70 CCAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGGCGCGGTACCGTGGGAC 129  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 130 GCAGCAGCTCAAGTGGATCCGCTCAACGACTTATTCGAGCGCGCAGCATGAACATT 189  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 190 ATGGCACCGAGTTCATCGCTGGACCCCAACACACCCGACAGGCTGTACCTCGCGCAG 249  
Qy 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
Db 250 GGGCGCTATGTCGGCGACGAGTGG-----GCGCGGTCTTATGTGTCGAAGAC 297  
Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
Db 298 CGCGCCAGTCTTTTACAACTTACAGTGCCTGCTCCCGATCGCGCGCCACGACATGGGA 357  
Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
Db 358 CGCAACAATGGCGAGCGCTGCTGTCAACCGTTCACCTCGAACGAGGTCTGGATGGGT 417  
Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
Db 418 ACGCGTACA---GAGGTATCTGGAAGAGATTTCGACCGCGCCGACACCTGGACAAACGTC 474  
Qy 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
Db 475 ACGTCCATCCGCGACGCGTTC-----ACCAACGCGTATCGGATACACG 516  
Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
Db 517 TCG-----GTCAATTTTCGACCCC-----GAA 537  
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
Db 537 TCG-----GTCAATTTTCGACCCC-----GAA 537

538 CGTAATGGACCATCTACGGAGCGGAGTCCCGCCAGGCGC---ATGTACGTACGAC 594  
225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
595 GACGGCGGTCTCTCGTGGGAGCGAGTGGCGGCGCCAGCCGTCAGCTGGCTCAACAGGACC 654  
242 -----ProHisGlyVal 246  
655 ACGGGCGGTTCCCGGACAAGAGCCGCGTGCATCGCGCCGACGCCCATGAAGTCGCT 714  
247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
715 CTCACCCCG---AAGTCTCTACGTGACTTACGCGCGATACCTGGTCCATGGGCG 768  
267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285  
769 GTCACGTTCCGGCGAAGTCTGGCGCAGAACCGACCTCGGGCGCTGGGACGACATTACT 828  
286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
829 CCCCAGGTCGGAACACTCGTCCGCTGCCGCTACAAACACGACGCTTCCTCGGGCGGA 888  
300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
889 TTTTGGGTCTCAGCGTCCAGCGGACCAACCCACCCGCTCTGCTGTATCACC---CTC 945  
320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
946 GACCGCGACCCCGACCCGCTCGACGACATCTACCTCTCAACCGATCGCGCGGACCC 1005  
337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
1006 TGAAGAGGACGTACCCAGCTCTGCTCCGCTCCAACTCGAAGTAACTGGGGCCACCCG 1065  
348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
1066 ACTAACCGCGCGGTAC---AAGGACGCGACGCTGTTCCGTGGCTCGACTTCAACAAC 1122  
368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379  
1123 GTTCCCGAGTGGGGGGATACGGTGGCGCGACGGTACGCCCGCTCACCAGTTTGGC 1182  
380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerArgMetLeuTyrGlyThr 399  
1183 TGGTGGATGAGCGCTGTGTATCGATCGCTTCAACCCCGACGACCTGATGATCGGCACG 1242  
400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415  
1243 GGGGCGACCATCTGGCGGACCGACGCTCTCCCGTGTGCAAGAGACTGG----- 1293  
416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431  
1294 -----GCGCGAGCTGGTACCTCCAGATCGAGTATCGAGAGATGCG 1338  
432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
1339 ATCTGTCTCGCTCGCTCGCCCAAGAGCGCGCGCTCTCGTGGCATCGTGCACATT 1398  
452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
1399 ACGCGCATGAAGCAGCAGCAGCACTTCAACAAG---CCCCAGAAGATGTTGTGCGCCCCAG 1455  
472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
1456 TTCTCCAACTCGACAGCATCGAGCTGGGGCACTTCCCAACGTTGTGTCGCGGCC 1515  
492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
1516 GGATCTCGGACACAGTACGACACCGCGTGGCGCGCGTGGCTACGCGACTGACGCG 1575  
512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 526  
1576 GGAGAGCGGTGACCATCTTCCCTACCTGCCCTCTGGCATGAACGCGAGCCACTACCAG 1635

527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544  
1636 GGCGACCATGATTCAGTTCGACGCGGCGGACGATCGTGTGTCGACCAAGCTTGAC 1695  
545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
1696 GAGCAGGCTCGGACCGTGTACTCGCAGCATATGCAAGACGTGGTCT----- 1746  
565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578  
1747 ---GTTCCGCTGGCGACCTGAAGGCCGACACTGCCAATGTGTCTCGGACAGGTCAG 1803  
579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598  
1804 GATGGCAGCTTCTACCTACCGATGGCGCAAGTCTCTGCTCGACCGACGCGGGAAG 1863  
599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
1864 TCGTATGCCCGCAAGGCGCGGACTTGTCACT-----GGCACATGCTCATGCTGCC 1917  
619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
1918 GTGAACCTCTGGGTGGCGCGGACGCTGGGTGCTGTTCCCGAGGGCGTCTCTTCAC 1977  
638 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
1978 TCGACCGACTTGGCGCTCTGTTACGAGGGTAGTACCCCAACGCGACCTCGTGAGC 2037  
657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
2038 GTCCGCGCCCCCAAGTCCAAGTCGACGCAAGAGCTAGCGCGCTCCCGGCTTTC 2097  
673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690  
2098 ATCTGGGACCGACCAAGCTGGAAGCGCATCGGCTGTACCGTCCGACGACCAACGGC 2157  
691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
2158 AGCACCTGACGCGCTCAATGACGAGGACCACTACTCTCGGCG---CCACCATGATC 2214  
711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
2215 GAGGCGCGCCCAAGGCTTACGCGCGCTGTATCTAGGCACGAACGCGCGGTATCGTG 2274  
731 TyrGlyAspIle 734  
2275 TAGCGCGACCTT 2286

## RESULT 11

US-10-395-241-11  
; Sequence 11, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; PRIORITY FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 11  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Geotrichum sp. M128  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (120)..(2558)  
US-10-395-241-11









QY 85 ----- 85  
Db 662 ACCCGTAGCTCGGACCAACGCGAATAGCGGTCCTGCTAGTGTTCACATACGCGGTAC 721  
QY 86 GlyMetThrAsnSerTrpAsp----- 93  
Db 722 GGCATGTATACAAACAGCTGGTC-TGTGATGCTCTCAGATCTAGACCTATGATGGAGC 780  
QY 93 ----- 93  
Db 781 GCCGTACATATGCTGTGACACAGACACTACAGAGAGTCTAGATCTGGATACTAACTGC 840  
QY 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 841 CTGACATTGGCCATATATAGGATCCGAGTAATGAGCCATCATTTGCTCGTCAGACCGC 900  
QY 105 ----- 105  
Db 901 GACTGTAAACGGTATATATCCCTAGGCTCATTTACCTCGGTAGTAAGCGAGCAGCTGGCG 960  
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 961 GCGCAACAGTGTCTCTTCAACCAACTTGCCCTTCAAAGTCGGGGTAAACATGCCAGGACG 1020  
QY 125 ----- 125  
Db 1021 CCGCGTTGCACAGGAAGTGTGTGAACGGGAAGTTTCAGCCCCCATTTGTACGGTCTCTGCG 1080  
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 1081 GGAGCGGAGAGCGTCTGCTGTGCATCCGCCAACTCCCAACATCATCTACTTTGTGTCT 1140  
QY 146 Pro----- 146  
Db 1141 CCTCGGCTCTCGCAGCCGACAGCTAGCGCGTTGAGTGTAGTAGATGAACCAACA 1200  
QY 147 SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164  
Db 1201 CGCTCAGGAACGGCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCG 1260  
QY 165 -----Thr 165  
Db 1261 GCGAGTCTTTCGGGAGACCTTCAGATGCTCGCGCGCAGCTGGAAGTTCCAGAGC 1320  
QY 166 AsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGlyTyrGln--- 184  
Db 1321 TCGTTACGGCAACTGGGACGTACATCCAGACCCGAGTGATCCAACGGCTACACAGC 1380  
QY 185 -----Ser 185  
Db 1381 AGCAAGTGCCTGTGACCTCGCATGTAGGTCTGGGCTCACTAAGGTTCGCGATGTTGTCG 1440  
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 1441 GACAAAGCAAGGACTCATGTGGGTTCAGTTCCAGTCAACACGAGCAGCAGCGGGGAGCC 1500  
QY 205 ----- 205  
Db 1501 CTGTTGTTCTCTGAGTACACCAACATGCAAGCTGAGTGTGCTGCTGGCGCCCTCGC 1560  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 222  
Db 1561 ACGTCTGTATCTTTGTTGGCAGCGCTGATACATCACTGCTTCAGTCTATGTGAGCAGC 1620  
QY 223 -----SerArg 224  
Db 1621 TGCAGAGCATGAACAAACCGTGGCGACTATTTGTAGTGACGAAGTCAGATACACTCGTGC 1680  
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis--- 243  
Db 1681 AATGCCGCTCCACGTGGAGTGTGTACCGGGGACCGAGGGAATATCTTCTCTCACAG 1740  
QY 244 -----LysGlyValPhe 247

Db 1741 TTACGCCCGAGGTGCACCTCACGACATGGCCCCGTCGGTCCCTTTATGAAAGGAGTGTTC 1800  
QY 248 -----AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
Db 1801 GCGAAACTGCAGCCACGACGAGAGGCTTGTATCTGACCTATTCCGATGCGACAGGCGCG 1860  
QY 264 ----- 264  
Db 1861 CGCTTTGAGCTCGGTCTCTCTCCGGAACATAGATGGATAAGGCTACCGTGTCCCGGC 1920  
QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284  
Db 1921 TATGATGGCACACTTGGCTCAGTGTGGAGGTACACATTCAGGGGGAACCTTGGAAAGAC 1980  
QY 285 IleSerProValProSerThrAspThr----- 293  
Db 1981 ATACTACCGTGTGAACCGAGTCACACCTCCATGCTGTAAAGTCCCTTGAACCTTCTG 2040  
QY 294 -----AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp--- 306  
Db 2041 ATCACCCCTGTCTCTGGATCAGATCTATCTTTGGCTTTGGCGCCTTGGCCTCGATTG 2100  
QY 307 -----Arg 307  
Db 2101 TAGTGGGACAGAGACCTAGTCTAGATATGAACCGGAACCCGCGAACCAGGACTAAAC 2160  
QY 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr----- 325  
Db 2161 CAAAGCCAGGAACCTTGTGTGTCTTCTTGAACCTTTGTGGCCAGATGCTCAGCTG 2220  
QY 326 -----IleIle 327  
Db 2221 GTTTTCGGTCTTGGGAACAACAAGAACTTGAGAACCACCGCTCTACGAGTCGAC 2280  
QY 328 PheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrPro 347  
Db 2281 TTTTCGGTCCAGCCAGCTCTCGGACACATCGAGCCCGATCTGGCGGTGGCGAGTATCCG 2340  
QY 348 AsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro----- 361  
Db 2341 AAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGGCTAGACCCGCAACCCGCTCGATAGG 2399  
QY 362 -----Trp 362  
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTCTTAACGATCCGATCGAATGG 2459  
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370  
Db 2460 CTGACTCTGGATAATGATGTGTAGAGTCACCTCAGTGAGAAITGCTAGGTACGCTTACC 2519  
QY 371 -----ProProValPro----- 374  
Db 2520 GACTGACTGTCTTCAAGACTCCCAAGCACCGGTGATCAAGAACACTTTTATCGATGTG 2579  
QY 374 ----- 374  
Db 2580 ATGACTGACACGAAGTTCGTGAGGTTTCGTGGCACCTAGTCTTGTGTAATAGCTACAC 2639  
QY 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385  
Db 2640 TCGAGCGAGTCACCGTCCGATGCTCTCATCAAGCGCTCGGCTGGATGATTGAGTCTCTC 2699  
QY 386 Ala----- 386  
Db 2700 GGCTCGCTCAGTGGCAGGCTACCGAGTAGTTCGCGAGCCGACCTACTAACTCAGAGAG 2759  
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 2760 CAGATTGACCCCAACCAGACGACCAACCTACCTGGCTCTACGGCACCGGAATGACAACTTTGGC 2819  
QY 406 Thr----- 406

Db 2820 GTCTAACTGGTGGTGTGTTGGTGACCGAGATGCGGTGGCCCTTACTGTTAGAAACCG 2879  
Qy 407 ---AsnAspLeuThrLysPheSerGlyGlnIleHisIleAlaProMetVallys 425  
Db 2880 CGCCAGATCTCACCACTGGGACAGCGCCACAATGTGTCAATCAATCACTGGCAGAC 2939  
Qy 426 Gly 426  
Db 2940 CGGTGTCTAGAGTGGTGGACCTGTGCGCGGTGTGTACAGTTAGTTAGTACCGTCTG 2999  
Qy 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProIleuile 445  
Db 3000 CGCATCGAGGAATCTCGTCCAGAGACCTGGCTCTGCACCCGCGGAAGCAGCTATTG 3059  
Qy 446 Ser 446  
Db 3060 CGGTAGTCTCTTAAGAGCAGGTCTTGACCGGAGACGTGGCGCGCTTTCGCTCGATAAC 3119  
Qy 447 ---AlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaVal 462  
Db 3120 CCCGAGTGGAGACGACACAGCGCTTACCTTTGCCAGCAGAAACGACCTCGGACATCG 3179  
Qy 462 462  
Db 3180 CGCGTCTGACCTCTGCTGTGCCGAAGTGGAAACGGTCTGTTGCTGGACCCCTGTAGC 3239  
Qy 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 481  
Db 3240 CGCAGACGGTCTGGGCAACCCACATCGGGCCACCTCGACAGCGTCTGACTACGCCGG 3299  
Qy 482 482  
Db 3300 AGCGTCTGCCAGACCGGTGGGTGTACCGGTGGAGCTGCTCGACGTGATGGGCC 3359  
Qy 483 LeuAsnProSerIleValAlaGlySer 493  
Db 3360 TACTCGTCAAGAGCGTCTCGCGTCCGCAACACCGCGGCAACAGGTGGCCATC 3419  
Qy 494 ---PheAspProSerSerGlnProAsnAspArgHisValAlaPhe 507  
Db 3420 TTGAGCCAGTCTCGACAGCGCGGACCGGTGTGGCGCGCGTGTGTCACCGGTAG 3479  
Qy 508 SerThrAspGlyGlyLysAsnThrPheGlnGlySerGluProGlyGlyValThrThrGly 527  
Db 3480 AGTCCGACGGCGCGACGTGAGCATCGACTAGCGCGCGACACGTCCATGACCGC 3539  
Qy 528 Gly 528  
Db 3540 GCGAGGTGCGCGCGCTGACCTCGTAGTGTGATGCGCGCGCTGTGCGAGTACTTGGC 3599  
Qy 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValThrAlaProGlyAspProGly 547  
Db 3600 CGCAGGTGGCTATTTCGCGCGCGCGACGATCCTCTGTGTCAGCGCTCGTTCGCGC 3659  
Qy 548 548  
Db 3660 GGTGCCACCGGTAAGCCGCTGCTGTCTAGGAGACGAGTGGCGGAGCGGCG 3719  
Qy 550 ValValThrAlaValGlyPheGlyAsnSerThrAlaAlaSerGlnGlyValProAlaAsn 569  
Db 3720 CTGCGAGGCTCG---CAGTTCCAGGGCAGCTTTCCTCGCTCGAGCTGCGCGCGC 3776  
Qy 570 AlaGln 571  
Db 3777 GCGCAGCTCGGAGGCTCAAGGTCCCGTCCGAACGAGGAGAGAGTCCGAGCGGCGCCG 3836  
Qy 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589  
Db 3837 CGGCTCATCGCTCGGACAGAAAGACCAACAGCGTCTTCTAGCGCGGTCTCGATCGACC 3896  
Qy 590 Phe 590  
Db 3897 TTTTAGTAGCGGAGCTGTTCTTCTGTTGTGCGCAGAGATGCGGCGGAGGCTAGTGG 3956

Qy 591 ---TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609  
Db 3957 AAAACAGCTCAGCAAGGACACCGGACAGCTTC-----ACGCGCGG-----CCCAAG 4004  
Qy 610 SerGlyAlaValGlyValMet----- 616  
Db 4005 CTGGGACGCGAGGATGAGTGTCTGTGGCGGTGTGCGAGTGTGCGCGCGCGGTTTC 4064  
Qy 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627  
Db 4065 GACCGTCTCGTCCCTCGATCGGGATATCGTCTCACCCGACACCGCGGCGCAGTTG 4124  
Qy 628 TrpLeuAlaAlaSerSer----- 633  
Db 4125 TATGTCTCGACCGAGGTAGGCCCTATAGCGACGAGTGGGTGGTGGCGCGTGAAC 4184  
Qy 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644  
Db 4185 ATACAGAGTGTGCTGTCTCGCATATTCGCTCCACAGACTCGGGCAGACCTTTGGCCAA 4244  
Qy 644 ----- 644  
Db 4245 GTCTCCACCGCCTGAAGCGGTATAGGCGAGGTGTCTGAGCCGCTGTGGAACCGTTC 4304  
Qy 645 -----TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662  
Db 4305 CAGAGTGGCGGACTCCAAACACCTACAGATCGCCCTGGGTGGGTCTCAGGCTCGAAC 4364  
Qy 663 -----AlaProGlySerSerTyr 668  
Db 4365 TGGAACTGTATGCTGTGGTGTGGATGTCTAGCGGACCCACCGAGTCCGAGCTTG 4424  
Qy 669 ProAlaValPheValGlyThrIleGlyValThrGlyAla-----TyrArgSer 686  
Db 4425 ACCTTGGACATACGATCGGCACC-----GGCCGCTAGGGGCTCGCTCTACGCCAGT 4478  
Qy 687 AspAspCysGlyThrThrTrpVal----- 694  
Db 4479 GGAGACAGCGGCTCTCTGAGCGGTGCGCGGCGAGTCCCGGAGGAGATGCGGTCA 4538  
Qy 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly----- 704  
Db 4539 CCTCTGTGCGCGGAGGACCTCGGACATCCAGGGCTCCAGGGCTTCGCTCCATCGAC 4598  
Qy 704 ----- 704  
Db 4599 AGCACAAGTGTGCGCGAGCGCGCTGTAGTCCGAGGTCCTCCGAAGCGAGGTAGCTG 4658  
Qy 705 AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThr 724  
Db 4659 TCGTGTTCAGCGCGCTGCGGCGAGCACCGCGCGG-----CAAGTCTACGTGGGACCC 4712  
Qy 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 4713 AACGCGCGGGGCTCTTTACGCTCAGGCGTGTGCGCGGCC 4754

## RESULT 14

US-10-282-122A-30151  
; Sequence 30151, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.

```

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30151
; LENGTH: 7407
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30151

Alignment Scores:
Pred. No.: 1,27E-12 Length: 7407
Score: 238.00 Matches: 205
Percent Similarity: 32.59% Conservative: 88
Best Local Similarity: 22.80% Mismatches: 314
Query Match: 5.90% Indels: 292
DB: 13 Gaps: 50

US-09-917-376-3 (1-740) x US-10-282-122A-30151 (1-7407)
QY 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
Db 3250 GTCAACCTGAGCAATGGCAGCAGCTCAGCGC-----ACCGCCGAGCGCGC 3297
QY 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTyrPheAlaAlaAsnGlyArg 50
Db 3298 AGCAGCGGTGATTCTCACCAGC-----GGCAACGGCAAT 3330
QY 51 TrpIleProLeuLeuAspTrpValGlyTyrPheAsnThrGlyTyr----- 65
Db 3331 CCGATCGCCGAGGTACACCGCAGCGCAGCGCACTGACCTACACCCCGCTCCACGCCG 3390
QY 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrLysValTyr 82
Db 3391 ATCGCCACCGGTACTGTGTGTCAAGGTGTGGCCAGCAGCCCTCCGGTAAACAGCAGCCG 3450
QY 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96
Db 3451 CCGCGCAGCGTGACCGTCTGATTCAGCGCGCGCGCGCGCGGTGATCAACCGAGCAAC 3510
QY 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115
Db 3511 GCGGTGTCATCAGCGGCAACCGCGAGCGCGGTGCGCAGCAGCTCAGCGTCAACCGTCA 3561
QY 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135
Db 3562 ----GATCGCGCGCGCAACCGATAGGCGAGGTACCGCGCGC----- 3600
QY 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThr 155

```

```

Db 3601 -----GGCAGCGCAACTGGAGCTTCCAG 3624
QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyIleThrTyrIleAla 175
Db 3625 CCG-----GGACCCCGCTGGCC 3642
QY 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185
Db 3643 AACGGCAGCGTGATCGTCGCCACCGCCAGCCGCGCAATACCGCGCCGCGCAGGCC 3702
QY 186 -----AspIleGlnGlyValTyrValAlaPheAspLysSerSerSer 201
Db 3703 GCCACCACCGTGGACGCGGTGGCGCGCGCGGTGATCGATCCGAGCAACGGCAGC 3762
QY 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218
Db 3763 ACCATCAGCGGACCGCGGAGCGCGGAGGTGATCTCACCAGCGGCAACGGCAAC 3822
QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237
Db 3823 CCGATCGGCAAAACCCAGCGCGCAGCGCAACTGGAGCTTCACGCCCGCGCAGCGC 3882
QY 238 -----ThrGly--- 239
Db 3883 CTGGCCAACCGCAGCGGTGTCAACCGCGTGGCCCGCAGGACCTCGCGGCAATACCGGCCG 3942
QY 240 -----PheIleProHisLysGlyValPheAspProValAsn 251
Db 3943 CAGGGCAGCATTACCGTGGACGGTGGCGCGGCAACAGCGCTGTGGTCAATCCGAGCAAC 4002
QY 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268
Db 4003 GGCAACCTGCTCAACCGTACCGCGCGCGCAGCAGCGCTTGACCTTGACCGCGCAAC 4062
QY 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285
Db 4063 GGCAACCCGATCGCGCAGCAGCAGCGCGATGGCAGCGCAACTGGAGCTTCACGCCCGC 4122
QY 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301
Db 4123 TCGCAACTGCCCAACCGCAGCGCGGTGTCAACGTGACCGCGCAGCGCGCGCGCAATACC 4182
QY 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315
Db 4183 AGCTTCCCGCTACACGACGAGGTGATTCCTCGTCGCGTGGATCCCGCAGGTGGATCCG 4242
QY 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330
Db 4243 AGCAACGGTTCGGTGTATCAGCGCAGCGCGCGCAACCACTCATCATC----- 4296
QY 331 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAsp 342
Db 4297 ACCGATGGCAACGGCAACCGGATTCGCCAGGTACCGCGCGCGCGCAACTGGTCC 4356
QY 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354
Db 4357 TTCACCTCAGGATCCCGTGGCGGATGGACCGGTGTCAACGTGGTGGCGCGCGCA 4416
QY 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnPro 372
Db 4417 AGCAATGTGCACAGTGGCGCGGTGATCCTGTGTGGTGGCGCGCGCGCGCGCGCG 4476
QY 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392
Db 4477 GTG-----ATCGATCCGAGCAACCGC 4497
QY 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409
Db 4498 ACGGAG---ATAAGCGGTACCGCGGAGCGCGCGCGGTGATCTCTCACC----- 4545
QY 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGlu 429

```

Db 4546 -----GATGGGGGGGCAAC-----CCGATCGCGCCAGGCCACCGCCGAC 4584  
Qy 430 ThrAlaValAsnAspLeuLeuSerPro-----ProSerGlyAlaProLeuLeu 445  
Db 4595 GCGCGGCAACTGGACGTTTCAACCCGAGCACCCTGGCCACCGGACCGGTGATCAAC 4644  
Qy 446 SerAlaLeuGlyAspLeuGlyPheThrHis-----AlaAspVal 459  
Db 4645 GCGGTGGCCAGGACCGCGCGCAATACCAAGCGGTCCGGCCAGCGTCAACCGTGCATGCC 4704  
Qy 460 ThrAlaValProSerThrIlePhe-----ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCCCGCGCGCGCGGTGATCAATCCGAGCAATGAGTGCATCAGCGGTACG--- 4761  
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 -----GGGAAGCGGGGCGACCGGTGATCTCCAGCGGCAACCGGACCGGATC 4812  
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GGCAG-----GTACCGCGGACGCGGCAACTGGAGCTTC 4851  
Qy 516 ----- 516  
Db 4852 AGCCCGGACGCGCGTGGCCACCGGTGCGTGTGATCAATGCGTGGCCCGGACGCGCC 4911  
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525  
Db 4912 GGCACACACAGTCCACCGGCGCCACCGTGCAGTCTGCGGCGCCAGCGCCCGGTG 4971  
Qy 526 -----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGACGACGAGTAGGTGATCGCGGTACCGCGAGGTGGGCCAGGTGATC 5031  
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGACGCAACCGGCAACCGGACCGGTGATCGCGGCGGACCGGCGGCAAC--- 5088  
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579  
Db 5089 TGGAGTTTCAGCGCGGACGCGC----- 5112  
Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg----- 592  
Db 5113 -----CTGTCAATGGACGCGTGTGATGCGTGGTGGCCGACGCGCTGCC 5157  
Qy 593 ---SerThrAspGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCACACCGCGCGCGGTGATCGACACCGGTGGACGCGGTGGCCCGCCACCGCGGTG 5217  
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626  
Db 5218 ATCGACCGGACGAGTGTGCACTACGCGGCAACCGCGGCGGCGTCCGGGTGATC 5277  
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645  
Db 5278 CTCACCGATGTCATGCAATCCGATCGCGCAGACCTTGGCGGCGGACGCGGCACTGG 5337  
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661  
Db 5338 AGTTTCAGCGCGGCGCGCGTGGCCACCGGACGCGGTGATCAATGCGTGGCCGAGGAC 5397  
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 681  
Db 5398 CCGGCGCGCAATACCGCGCGCGGCGGACCGGCGGACCGGTGGACCGGTGGCTCCGCGCCACG 5457  
Qy 682 GlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuLeuAsnAspGlnHis 701  
Db 5458 -----CCGGTGTATCAATCCGAGCAAC--- 5478  
Qy 702 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 721  
Db 5479 -----GGCAGGTGATCACCAGGACC---GCCAGGTGCGGCGCCAGGTG 5520

Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 5521 ATCTCTACCGATGGCAACGGCAACCGATCGCGGACGACACCGCCGCGCAGTGGT 5577

## RESULT 15

US-10-246-330-3  
; Sequence 3, Application US/10246330  
; Publication No. US20030166030A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole, George A.  
; APPLICANT: Mah, Thien-Fah  
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF  
; FILE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE  
; FILE REFERENCE: 14537-002001  
; CURRENT APPLICATION NUMBER: US/10/246,330  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: US 60/323,241  
; PRIOR FILING DATE: 2001-09-18  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 7407  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1).....(7404)  
US-10-246-330-3

## Alignment Scores:

Pred. No.:	1,27e-12	Length:	7407
Score:	238.00	Matches:	205
Percent Similarity:	32.53%	Conservative:	88
Best Local Similarity:	22.80%	Mismatches:	314
Query Match:	5.90%	Indels:	292
DB:	15	Gaps:	50

US-09-917-376-3 (1-740) x US-10-246-330-3 (1-7407)

Qy 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30  
Db 3250 GTCAACCTGAGCAATGGCAGCAGCTCCTACGCGC-----ACCGCGGACCGCGC 3297  
Qy 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
Db 3298 AGCAGCGTGTCTCACCAGC-----GGCAACGGCAAT 3330  
Qy 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65  
Db 3331 CCGATCGCGAGGTCAACCGCCGAGCGGCGGCACTGACCTACACCGGTCCACGCGC 3390  
Qy 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82  
Db 3391 ATCGCAACCGGTACTGTGTCAACGTGGTGCCAGGACGCTCCGCTAACAGCAGCGCG 3450  
Qy 83 AlaAlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAsp 96  
Db 3451 CCGGCGAGCGGTGACCGTCAATTCAGCGCGCGCGCGCGGTGATCAACCGGAGCAAC 3510  
Qy 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
Db 3511 GCGTGTGTCATCAGCGGCGACCGCGGCGGTGGCCAGGTGACCTCACC----- 3561  
Qy 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135  
Db 3562 ---GATGCGCGGCGCAACCCGATAGGCGAGGTCAACCGCGC----- 3600  
Qy 136 AsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
Db 3601 -----GGCAGCGGCACTGGAGCTTCACG 3624  
Qy 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175

Db 3625 CCG-----||||| :|||  
Qy 176 Asn-----ProThrAspThrThrGlyTyrGlnSer 185  
Db 3643 AACGGCAGGTGATCGTCGCACCGCCACCGACCGCAATACCGCGCCGAGGCC 3702  
Qy 186 -----AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201  
Db 3703 GCCACACGGTGGAGCGGTGGCGCGCGCGCGGTGATCGATCCGAGCAACGGCACG 3762  
Qy 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
Db 3763 ACCATCAGCGGACCGCGGAGCGCGGGCCAAAGGTGATCTCACCGCAGCGCAACGGCAAC 3822  
Qy 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
Db 3823 CCGATCGCGGAACACCGCGGACCGGACCGGCACTGGAGCTTCACCGCGGACGCGG 3882  
Qy 238 -----ThrGly--- 239  
Db 3883 CTGGCCAAACGGCAACGGTGGTCAACCGCGTGGCCAGGACCTCGCGGCAATACCGGCGCG 3942  
Qy 240 -----PheIleProHisLysGlyValPheAspProValAsn 251  
Db 3943 CAGGGCAGCACTACCGTGGACCGCGGTGGCGCGCAACACCGCTGTGTCAATCCGAGCAAC 4002  
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 268  
Db 4003 GGCAACCTGCTCAACGGTACCGCGGACCGGGGACCGCGTGCCTTACCGCAGCGCAAC 4062  
Qy 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285  
Db 4063 GGCAACCGGATCGGCGCAGCACCGCCGATGCGACGGCACTGGAGTTCAGCGCCGCGC 4122  
Qy 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
Db 4123 TCGCAACTGCCCAACGGCACCGTGTCAACGTGTACCGCGGACGCGCGCGCAATACC 4182  
Qy 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315  
Db 4183 AGCTTCCCGTACACAGCGGTGATTCCTCGTGGCGGTGATCCCGAGGTGATCCG 4242  
Qy 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330  
Db 4243 AGCAACGGTTCGGTGTATCAGCGGCACCGCGGAGCGCGCAACACCATCATCATC- 4296  
Qy 331 ThrAspGlyGly-----AlaThrTrpTrpArgIleTrpAsp 342  
Db 4297 ACCGATGGCAACGGCAACCCGATTGGCCAGGTTCACCGCGCAGCGCGCAACTGTGTCC 4356  
Qy 343 TrpThr-----SerTyrProAsnArgSerLeuArgTyrVal 354  
Db 4357 TTCCTCAGGATCCCGCTCCGATGGCAGCGGTGTCACCGTGGTGGCGCGCAGCCCA 4416  
Qy 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnPro 372  
Db 4417 AGCAATGTCGACGTGCGCGCGGTGATCATCTGTGGATGGCGCGCGCGCGCGCG 4476  
Qy 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
Db 4477 GTG-----ATCGATCCGAGCAACGGC 4497  
Qy 393 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 409  
Db 4498 ACCGAG---ATAAGCGGTACCGCGGAGCGCGCGCGGTGATCCTCACC----- 4545  
Qy 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGlu 429  
Db 4546 -----GATGGCGCGGCAAC-----CCGATCGGCGCAGGCGCACCGCCGAC 4584  
Qy 430 ThrAlaValAsnAspLeuIleSerPro-----ProSerGlyAlaProLeuIle 445  
Db 4551 ATCCTCAGCGATGGCAACCGCAACCGGATCGGCGGAGACCAACCGCCCGCGCAGTGGT 5577

Db 4585 GGCAGGGGCAACTGGAGCTTACCCCGAGCACCCTCGGTGGCCAAACGGCAGCGTATCAAC 4644  
Qy 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHis-----AlaAspVal 459  
Db 4645 GCGGTGGCCAGGACCCGCGCGGCAATACCGAGCGTCCGCGCAGCGTCACCGTCGATGCC 4704  
Qy 460 ThrAlaValProSerThrIlePhe-----ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCGCCCGCGCGCGGTGATCAATCGAGCAATAGGATCGTCATCAGCGGTAC- 4761  
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 -----GCGGAAGCGCGGCGCAGGTGATCTCACCGACGGCAACCGCACCCGATC 4812  
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GGCCAG-----GTCACCGCGCAGCGGACGGCAGCGCACTGGAGTTC 4851  
Qy 516 ----- 516  
Db 4852 ACGCCCGGCAACCGCTGGCCAAACGGTGGTGTGATCAATCGCTGGCCAGGACCGCGCC 4911  
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525  
Db 4912 GGCAACAAACAGCAGTCCACACGCGCCCGTCGACTCGCTGGCGCAGCAGCGCCCGGTG 4971  
Qy 526 -----ThrGlyGlyThrValAlaAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGAGCAACGGTAGCTGCGGTGACCGCGGAGTGTGGTGGTGGTGGTGGTGGT 5031  
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGCGGCAACCGGCAACCGATCGCGCAGGTCAACCGCGTGGCAGCGGCAAC- 5088  
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerArgValAlaAsnPro 579  
Db 5089 TGGAGCTTCACCGCGGCGCGCG- 5112  
Qy 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg----- 592  
Db 5113 -----CTGTCCAATGGCAGCGGTGTCATCGGTGGCCAGGAGGTGGCC 5157  
Qy 593 ---SerThrAspGlyGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCAACACCGCGCGCGGTGACGACCAACCGTGGCGCGGTGGCGCGCGCGCGCGGTG 5217  
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626  
Db 5218 ATCGACCCGAGCAACGGTGTGCAACTCAGCGGCGCAGCGCGCAACCGCGGTGGTGGTGGT 5277  
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645  
Db 5278 CTCACCGATGGCAATGGCAATCCGATCGCGCAGACCTTGGCCAGCGCAGCGCAACTGG 5337  
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661  
Db 5338 AGCTTCACCGCGGCGCGCGGTGGCCACCGGCAACCGGTCGTCATCGCGTGGCCAGGAC 5397  
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 681  
Db 5398 CCGGCGCGCAATACCGCGCGCGCGCGCAGCACCACCGCGTGGACACCGTGGTGGTGGTGGT 5457  
Qy 682 GlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHis 701  
Db 5458 -----CCGGTGTGATCAATCCAGCAAC- 5478  
Qy 702 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyr 721  
Db 5479 -----GGCAGCGTGTATCAGCGGAC- 5520  
Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 5521 ATCCTCAGCGATGGCAACCGCAACCGGATCGGCGGAGACCAACCGCCCGCGCAGTGGT 5577

Search completed: May 12, 2004, 18:05:33  
Job time : 11480.2 secs



Result No.	Score	Query Match	Length	DB	ID	Description
1	191.	40.6	1711	2	T31337	1,4-beta-glucanase
2	190.5	40.5	1779	2	T31085	xylanase - Caldice
3	190	40.4	508	2	G6593	cellulase (EC 3.2.
4	190	40.4	508	2	A26874	cellulase (EC 3.2.
5	188	40.0	499	2	JN0111	cellulase (EC 3.2.
6	186.5	39.7	1331	2	A48954	mannan endo-1,4-be
7	186.5	39.7	1742	2	T17120	cellulase (EC 3.2.
8	186	39.6	499	2	A27198	cellulase (EC 3.2.
9	182.5	38.8	915	2	A43802	cellulase (EC 3.2.
10	182.5	38.8	1039	2	S02711	cellulase (EC 3.2.
11	180.5	38.4	145	2	A41897	cellulase homolog
12	180	38.3	486	2	I40548	bifunctional cellul
13	172	36.6	700	2	B41897	cellulase (EC 3.2.
14	165.5	35.2	879	2	A47704	endoglucanase I (E
15	165	35.1	1854	2	S36859	cipA protein - Clo
16	149.5	31.8	505	2	S39962	endoglucanase - Er
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2.
18	141.5	30.1	986	2	S12021	thermoactive cellul
19	130.5	27.8	586	2	PC6006	scaffolding protei
20	130.5	27.8	1483	2	C97012	probably celluloso
21	126.5	26.9	1162	2	T30433	scaffolding protei
22	117	24.9	1230	2	S47466	cellulose 1,4-beta
23	114	24.3	1848	2	A44140	cellulose-binding
24	75.5	16.1	618	2	T08685	hypothetical prote
25	73.5	15.6	547	2	T25478	hypothetical prote
26	73.5	15.6	1428	2	AC2224	hypothetical prote
27	70	14.9	5188	2	B85547	probable RTX fami
28	70	14.9	5291	2	F06096	hypothetical prote
29	68.5	14.6	574	2	A69196	cell surface glycop



Db 419 QMGGNLTHTKFTVTLHKPKQGADTYLE 444  
 RESULT 5  
 JN0111  
 cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)  
 N:Alternate names: endo-1,4-beta-glucanase  
 C:Species: Bacillus subtilis  
 C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 28-May-1999  
 C:Accession: JN0111  
 R:Park, S.H.; Kim, H.K.; Pack, M.Y.  
 Agric. Biol. Chem. 55, 441-448, 1991  
 A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE616  
 A:Reference number: JN0111; MUID:91299280; PMID:1368694  
 A:Accession: JN0111  
 A:Molecule type: DNA  
 A:Residues: 1-499 <PAR>  
 A:CROSS-references: GB:D01057; NID:G216387; PIDN:BAA00859.1; PID:d1001323; PID:G216388  
 A:Note: the authors translated the codon ATA for residue 102 as Tyr  
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal processing  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose  
 A:Pathway: cellulose degradation  
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 Query Match 40.0%; Score 188; DB 2; Length 499;  
 Best Local Similarity 40.0%; Pred. No. 6.7e-12;  
 Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
 QY 4 GVKVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVYNCWAA 63  
 Db 353 GISVOYRAGDGMNSNQIRPOLQKNNNTVLDKDVARYWYNNKQGNV--DCDYAQ 410  
 QY 64 MCGGNIRASFGSVNPTPTADTYLQ 88  
 Db 411 LGCGNVTYKFTVTLHKPKQGADTYLE 435  
 RESULT 6  
 A48954  
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum  
 N:Alternate names: beta-mannanase  
 C:Species: Caldocellum saccharolyticum  
 C>Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
 C:Accession: A48954; B43745  
 R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.  
 Appl. Environ. Microbiol. 58, 3864-3867, 1992  
 A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain  
 A:Reference number: A48954; MUID:93119139; PMID:1476429  
 A:Accession: A48954  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1331 <GTB>  
 A:CROSS-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291  
 A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIIP:121577)  
 R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
 Appl. Environ. Microbiol. 57, 694-700, 1991  
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding  
 A:Reference number: A43745; MUID:91247819; PMID:2039230  
 A:Accession: B43745  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-337, 'PPROHOHQ' <LUE>  
 A:CROSS-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294  
 A:Note: the authors translated the codon CAC for residue 262 as Glu  
 A:Note: this sequence has been revised in reference A48954  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 Query Match 39.7%; Score 186.5; DB 2; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 2.8e-11;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVYNCW 61  
 Db 364 SQGIKVLVYANKETNTTIRPWLKVNSSGSSSIDLSRVTIRYWTYDGERAQSAIS-DW 422  
 QY 62 AAMGCGNIRASFGSVNPTPTADTYLQ 88  
 Db 423 AQIGASNVTFKFKLSSSVSGADYYLE 449

## RESULT 7

T17120  
 cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum  
 C:Species: Caldocellum saccharolyticum  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Nov-2000  
 C:Accession: T17120; A43745  
 R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.  
 Appl. Microbiol. Biotechnol. 43, 291-296, 1995  
 A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophilic  
 A:Reference number: Z18698; MUID:95336703; PMID:7612247  
 A:Accession: T17120  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1742 <TEO>  
 A:CROSS-references: EMBL:L32742; NID:G537499; PID:G537500; PIDN:AAA91086.1  
 R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
 Appl. Environ. Microbiol. 57, 694-700, 1991  
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding  
 A:Reference number: A43745; MUID:91247819; PMID:2039230  
 A:Accession: A43745  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1516-1544, 'A', 1546-1742 <LUE>  
 A:CROSS-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293  
 C:Genetics:  
 A:Gene: cellA  
 C:Keywords: glycosidase; hydrolase

Query Match 39.7%; Score 186.5; DB 2; Length 1742;  
 Best Local Similarity 42.5%; Pred. No. 3.7e-11;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSTLVYNCW 61  
 Db 704 SQGIKVLVYANKETNTTIRPWLKVNSSGSSSIDLSRVTIRYWTYDGERAQSAIS-DW 762  
 QY 62 AAMGCGNIRASFGSVNPTPTADTYLQ 88  
 Db 763 AQIGASNVTFKFKLSSSVSGADYYLE 789

## RESULT 8

A27198  
 cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IF03034)  
 N:Alternate names: endo-1,4-beta-glucanase  
 C:Species: Bacillus subtilis  
 C>Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 15-Oct-1999  
 C:Accession: A27198  
 R:Nakamura, A.; Uozumi, T.; Beppu, T.  
 Eur. J. Biochem. 164, 317-320, 1987  
 A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.  
 A:Reference number: A27198; MUID:87190397; PMID:3106035  
 A:Accession: A27198  
 A:Molecule type: DNA  
 A:Residues: 1-499 <NAK>  
 A:CROSS-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671  
 A:Experimental source: strain IF03034  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose  
 A:Pathway: cellulose degradation  
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation  
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 39.7%; Score 186.5; DB 2; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 2.8e-11;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Query Match 39.6%; Score 186; DB 2; Length 499;  
Best Local Similarity 40.7%; Pred. No. 1.1e-11;  
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;  
QY 4 GVKVQYKNDSPAGDNQIKPGQLQVNTGSSVDLSVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 353 GSVQYKAGDGVNSQNRQPQHKNGNATVDLKDVTARYWYNAKNGQN---FDCTYA 409  
QY 63 AMGCGNIRASFSGSVNPATPTADTYLQ 88  
Db 410 QIGCGNLTHKFTLHKPKGADTYLE 435  
RESULT 9  
A43802  
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Caldocellum saccharolyticum  
C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Jul-1998  
C:Accession: A43802  
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 56; 3117-3124, 1990  
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile  
A:Reference number: A43802; MUID:91136262; PMID:2126700  
A:Accession: A43802  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: DNA  
A:Residues: 1-915 <SAU>  
A:Cross-references: EMBL:X13602  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>  
Query Match 38.8%; Score 182.5; DB 2; Length 915;  
Best Local Similarity 41.9%; Pred. No. 4.8e-11;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
QY 3 GVKVQYKNDSPAGDNQIKPGQLQVNTGSSVDLSVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 349 GQIKVLYANKETNSTNTIRPWLKVNSSGSSIDLSRVITRYWYTVGGERAQSAYS-DWA 407  
QY 63 AMGCGNIRASFSGSVNPATPTADTYLQ 88  
Db 408 QIGASNVTFKFKLSSSVSGADTYLE 433  
RESULT 10  
S02711  
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum  
N:Alternate names: endo-1,4-beta-glucanase  
C:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)  
C:Species: Caldocellum saccharolyticum  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999  
C:Accession: S02711  
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.  
Nucleic Acids Res. 17, 439, 1989  
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exc  
A:Reference number: S02711; MUID:89098398; PMID:2789517  
A:Accession: S02711  
A:Molecule type: DNA  
A:Residues: 1-1039 <SAU>  
A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646  
C:Genetics:  
A:Gene: celB  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1039/Product: cellulase #status predicted <MAT>  
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>  
Query Match 38.8%; Score 182.5; DB 2; Length 1039;  
Best Local Similarity 41.9%; Pred. No. 5.5e-11;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
QY 3 GVKVQYKNDSPAGDNQIKPGQLQVNTGSSVDLSVTYVWF-TRDGGSSTLVNCDDWA 62  
Db 420 GQIKVLYANKETNSTNTIRPWLKVNSSGSSIDLSRVITRYWYTVGGERAQSAYS-DWA 478  
QY 63 AMGCGNIRASFSGSVNPATPTADTYLQ 88  
Db 479 QIGASNVTFKFKLSSSVSGADTYLE 504  
RESULT 11  
A41897  
cellulase homolog - Bacillus lautus (fragment)  
C:Species: Bacillus lautus  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 15-Oct-1999  
C:Accession: A41897; S27498  
R:Hansen, C.K.; Didarichsen, B.; Jorgensen, P.L.  
J. Bacteriol. 174, 3522-3531, 1992  
A:Title: celA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4  
A:Reference number: A41897; MUID:92276330; PMID:1592807  
A:Accession: A41897  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <HAN>  
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662  
A:Experimental source: PL236  
A:Note: sequence extracted from NCBI backbone (NCBIP:104604)  
Query Match 38.4%; Score 180.5; DB 2; Length 145;  
Best Local Similarity 43.9%; Pred. No. 1.1e-11;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;  
QY 7 VOYKNDSPAGDNQIKPGQLQVNTGSSVDLSVTYVWF-TRDGGSSTLVNCDDWAAMGC 66  
Db 1 LQRAADTNAADNQIKPFSFNKNGTSVAVLSTLKIRYFTKGSAAVNGW-IDWAQLGG 59  
QY 67 GNIRASFGSVNPATPTADTYLQ 88  
Db 60 SNIQISFG-NHRTGNSDTYVE 79  
RESULT 12  
I40548  
bifunctional cellulase precursor - Bacillus sp.  
C:Species: Bacillus sp.  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40548  
R:Han, S.J.; Yoo, Y.J.; Kang, H.S.  
J. Biol. Chem. 270, 26012-26019, 1995  
A:Title: Characterization of a bifunctional cellulase and its structural gene: the cel  
A:Reference number: I40548; MUID:96029707; PMID:7532793  
A:Accession: I40548  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-486 <RES>  
A:Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576  
C:Genetics:  
A:Gene: cel  
Query Match 38.3%; Score 180; DB 2; Length 486;  
Best Local Similarity 38.8%; Pred. No. 4.5e-11;  
Matches 33; Conservative 18; Mismatches 32; Indels 2; Gaps 1;  
QY 4 GVKVQYKNDSPAGDNQIKPGQLQVNTGSSVDLSVTYVWF-TRDGGSSTLVNCDDWA 63  
Db 353 GSVQYKAGDGVNSQNRQPQHKNGNATVDLKDVTARYWYNAKNGQN---DCTVEQ 410

```
Qy 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 411 LGCGNVSHVTTLHKPKQGADTYLE 435

RESULT 13
B41897
cellulase [EC 3.2.1.4] - Bacillus lautus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus lautus
C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999
C:Accession: B41897; S27499
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: ccla from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: B41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663
A:Experimental source: PL236
A:Note: sequence extracted from NCBI backbone (NCBIP:104605)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.6%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 4.5e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 59
Db 549 VNSDLVQYKDDNNATDNOIKHFNIQNGTSPVDLSSLTLYFYTKD-SSAAMNGWI 607

Qy 60 DWAMCGCNIIRASFGSVNPATPTADTYLQ 88
Db 608 DWAKLGGSNQISFGHNGA--DSDTYAE 634

RESULT 14
A47704
endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum
C:Species: Clostridium thermocellum
C>Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A47704
R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.
J. Gen. Microbiol. 139, 307-316, 1993
A:Title: Gene sequence and properties of CellI, a family E endoglucanase from Clostridium
A:Reference number: A47704; MUID:93171873; PMID:8436949
A:Accession: A47704
A>Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-879 <HAG>
A:Cross-references: GB:L04735; NID:g144807; PIDN:AAA20892.1; PID:g144808
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)
C:Keywords: glycosidase; hydrolase

Query Match 35.2%; Score 165.5; DB 2; Length 879;
Best Local Similarity 33.0%; Pred. No. 2.7e-09;
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 60
Db 736 IKGEVLQYANGAGATSNINPRPKIINNGTKAINLSDVKIRYYTYTKEGGASQNFW-CD 794

Qy 61 WAAMCGCNIIRASFGSVNPATPTADTYLQ 88
Db 795 WSSAGNNGVNTGNFFNLSPKEGADTCLE 822

RESULT 15
```

```
S36859
cipA protein - Clostridium thermocellum
N:Alternate names: probable cellulosome protein large chain SL
C:Species: Clostridium thermocellum
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S36859; S33527; S25767; S28659; T18261
R:Gerngross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: EMBL:L08665
R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615,1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that b
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615,1617-1768, 'R', 1770-1854
A:Cross-references: EMBL:X67406
R:Fujino, T.; Bequin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18261
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipa

Query Match 35.1%; Score 165; DB 2; Length 1854;
Best Local Similarity 36.1%; Pred. No. 6.9e-09;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKVOYKND-SAPGDNOIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSLTLYNC 60
Db 365 VSGNLKVFYNSNPSTTSINPQFKVTNTGSSAIDLKSLTRYTYTVDGQKQDTFW-CD 423

Qy 61 WAAMCGCNIIRASFGSVNPATPTADTYLQ 88
Db 424 HAAITGNSGNYGITSNVKGTFFVKMSSSTNNADTYLE 460

Search completed: May 11, 2004, 12:10:26
Job time : 6.17339 secs
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 2.42639 Seconds  
(without alignments)  
1909.933 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	40.4	499	1 GUN1_BACSU	P07983 bacillus su
2	190	40.4	499	1 GUN2_BACSU	P10475 bacillus su
3	188	40.0	499	1 GUN3_BACSU	P23549 bacillus su
4	186.5	39.7	1331	1 MANB_CALSA	P22533 caldocellum
5	186.5	39.7	1742	1 GUNA_CALSA	P22534 caldocellum
6	182.5	38.8	1039	1 GUNB_CALSA	P10474 c endogluc
7	180.5	38.4	145	1 YCEA_PAEIA	P29718 paenibacill
8	172	36.6	700	1 GUNA_PAEIA	P02934 clostridium
9	165.5	35.2	879	1 GUN1_CLOTM	Q01866 clostridium
10	165	35.1	772	1 CIPB_CLOTM	Q06851 clostridium
11	165	35.1	1853	1 CIPA_CLOTM	Q59394 erwinia car
12	149.5	31.8	444	1 GUNW_ERWCA	Q47096 erwinia car
13	149.5	31.8	505	1 GUNW_ERWCA	Q59395 erwinia car
14	147.5	31.4	504	1 GUNW_ERWCA	P50900 clostridium
15	147.5	31.4	914	1 GUX2_CLOS	P23659 clostridium
16	141.5	30.1	986	1 GUNZ_CLOS	P38058 clostridium
17	114	24.3	1848	1 CBPA_CLOCL	P01809 mus musculu
18	67.5	14.4	118	1 HV39_MOUSE	P53116 methanococc
19	67.5	14.4	215	1 FLAI_METVA	Q9fsg7 malus domes
20	66.5	14.1	246	1 TPIA_MALDO	P71553 m bifunctio
21	65	13.8	523	1 PUR9_MYCTU	Q8rdj9 thermonaer
22	64.5	13.7	556	1 LLVD_THETN	P32660 saccharomyc
23	64	13.6	1571	1 ATCS_YEAST	P57719 thermoplasm
24	63.5	13.5	247	1 FLAI_THEVO	O17446 schistosoma
25	63.5	13.5	465	1 TV3H_SCHNA	Q9upv9 homo sapien
26	63	13.4	953	1 O106_HUMAN	P15648 h peroxisom
27	63	13.4	1581	1 PRXB_HUMAN	P33693 rhizobium m
28	62.5	13.3	269	1 EXOK_RHIME	P29352 mus musculu
29	62.5	13.3	802	1 PTN8_MOUSE	Q99n50 mus musculu
30	62	13.2	950	1 STL2_MOUSE	P09815 pseudomonas
31	62	13.2	1210	1 ICEN_PSEFL	P50937 rhodobacter
32	61.5	13.1	252	1 HIS6_RHOSH	P38108 pseudomonas
33	61.5	13.1	316	1 MUCB_PSEAE	

34	61.5	13.1	328	1 HAIQ_MOUSE	P14428 mus musculu
35	61.5	13.1	368	1 HAIW_MOUSE	P03991 mus musculu
36	61.5	13.1	489	1 GATA_CAUCR	Q9a510 caulobacter
37	61.5	13.1	678	1 YIHQ_ECOLI	P32138 escherichia
38	61	13.0	120	1 HV5Q_MOUSE	P06329 mus musculu
39	61	13.0	338	1 RTCA_ECOL6	Q8fcs8 escherichia
40	61	13.0	338	1 RTCA_ECOLI	P46849 escherichia
41	61	13.0	342	1 RTCA_ECO57	P58127 escherichia
42	60.5	12.9	1157	1 C9CA_BACTO	P58733 bacillus th
43	60	12.8	312	1 LECF_ALEAU	P18891 aleuria aur
44	60	12.8	1226	1 PAT2_CAEEL	P34446 caenorhabdi
45	59.5	12.7	261	1 COTN_BACSU	P54507 bacillus su

ALIGNMENTS

RESULT 1  
GUN1\_BACSU STANDARD; PRT; 499 AA.  
ID GUN1\_BACSU  
AC P07983;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase).  
GN BGLC OR GLD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DLG;  
RX MEDLINE=87194581; PubMed=3106328;  
RA Robson L.M.; Chambliss G.H.;  
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";  
RL J. Bacteriol. 169:2017-2025(1987).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

EMBL; M16185; AAA22496.1; ALT\_INIT.  
PIR; A26874; A26874.  
HSSP; O85465; IA3H.  
InterPro; IPR001956; CBD 3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR001547; Glyco\_Hydro\_5.  
Pfam; PF00942; CBM\_3; 1.  
Pfam; PF00150; cellulase; 1.  
ProDom; PD001947; CBD 3; 1.  
ProSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 499 ENDOGLUCANASE.  
FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
SQ SEQUENCE 499 AA; 55187 MW; 339D04EB95A63BE1 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;  
Best Local Similarity 43.0%; Pred. No. 1.3e-13;  
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVTRYWF-TRDGGSTLYVNCDA 62

```
Db 353 GVSQYKAGDGRVNSQIRPQLHKNNGNATVDLKVDTARYWYVKNKGON---FDCDYA 409
QY 63 AMCGNIRASFGSVNPTATDTYLQ 88
Db 410 QMGCGNLTHKFTVTLHKPKQGAADTYLE 435

RESULT 2
GUN2_BACSU
ID GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 01-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN BGLC OR GLD OR EGLS OR BSU18130.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAP115.
RX MEDLINE=87066783; PubMed=30241130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moraneli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degrees region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter.";
RL Microbiology 142:3097-3101(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Colightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic C., Ruchelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
```

```
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumatein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; CAA82317.1; -
DR EMBL; X04689; CAA28392.1; -
DR EMBL; X67044; CAA47429.1; -
DR EMBL; Z73234; CAA97610.1; ALT INIT.
DR EMBL; Z99113; CAB13696.1; ALT_INIT.
DR PIR; G69593; G69593.
DR HSSP; O85465; 1A3H.
DR Subtilist; BG10437; bg1c.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 ENDOGLUCANASE.
FT ACT_SITE 257 257 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 350 499 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 283 283 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55287 MW; 8F73557F711B3AE2 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 13e-13;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSDAPGDNIQKPLQVNTGSSVLDSTVTYVYWF-TRDGGSTLVNCDWA 62
Db 353 GISVQYRAGDGMNSQIRPQLQKNGNTVDLKVDTARYWYVKNKGON---FDCDYA 409
QY 63 AMCGNIRASFGSVNPTATDTYLQ 88
Db 410 QMGCGNLTHKFTVTLHKPKQGAADTYLE 435

RESULT 3
GUN3_BACSU
ID GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)
```



DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-giucanase)  
 DE (Carboxymethyl-cellulase) (CMCANE) (Cellulase).  
 GN BGLC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BS616;  
 RX MEDLINE=91299280; PubMed=1368694;  
 RA Park S.H., Kim H.K., Park M.Y.;  
 RT "Characterization and structure of the cellulase gene of Bacillus  
 RT subtilis BSE616";  
 RL Agric. Biol. Chem. 55:441-448(1991).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; D01057; BAA00859.1; -;  
 DR PIR; JN0111; JN0111.  
 DR HSP; 085465; 1A3H.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 1.  
 DR Pfam; PF00150; cellulase; 1.  
 DR ProDom; PD001947; CBD\_3; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 KW SIGNAL 1 29  
 FT CHAIN 30 499 ENDOGLUCANASE.  
 FT ACT\_SITE 169 169 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
 SQ SEQUENCE 499 AA; 55169 MW; 2E821E3DB8BACA04 CRC64;  
 Query Match 40.0%; Score 188; DB 1; Length 499;  
 Best Local Similarity 40.0%; Pred. No. 2.2e-13;  
 Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
 OY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDMAA 63  
 Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:  
 353 GISVQYRAGSGMNSGNRQPKQIKNGNTTVDLKDVTARYWNAKKNQNV--DCDYAQ 410  
 OY 64 MCGNIRAFSGVNPATPTADTYLQ 88  
 Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:  
 411 LCGGNVYKFVTLHKPGOGADTYLE 435  
 RESULT 4  
 MANB\_CALSA STANDARD; PRT; 1331 AA.  
 AC P22533;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-  
 DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-  
 DE mananase); Endo-1,4-beta-giucanase (EC 3.2.1.4) (Cellulase)].  
 GN MANA.  
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;

OC Caldicellulosiruptor.  
 OX NCBI\_TaxID=44001;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93119139; PubMed=1476429;  
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;  
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a  
 RT multidomain enzyme";  
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).  
 RN [2]  
 RP SEQUENCE OF 1-346 FROM N.A.  
 RX MEDLINE=91247819; PubMed=2039230;  
 RA Luehli E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a  
 RT gene coding for a beta-mannanase from the extremely thermophilic  
 RT bacterium 'Caldocellum saccharolyticum'";  
 RL Appl. Environ. Microbiol. 57:694-700(1991).  
 CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT  
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH  
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic  
 CC linkages in mannans, galactomannans, glucomannans, and  
 CC galactoglucomannans.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees  
 CC Celsius.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; L01257; AAA71887.1; -;  
 DR EMBL; M36063; AAA72861.1; -;  
 DR PIR; A48954; A48954.  
 DR HSP; Q06851; INBC.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 2.  
 DR Pfam; PF00150; cellulase; 1.  
 DR ProDom; PD001947; CBD\_3; 2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Hydrolase; Glycosidase; Cellulose degradation; Signal;  
 KW Multifunctional enzyme.  
 FT SIGNAL 1 41  
 FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.  
 FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).  
 FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).  
 FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).  
 FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).  
 FT ACT\_SITE 162 162 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 FT CONFLICT 338 338 T -> P (IN REF. 2).  
 FT CONFLICT 340 346 TPPTPT -> RQHQHQ (IN REF. 2).  
 SQ SEQUENCE 1331 AA; 146892 MW; FFC5A51BB8D8F0E0 CRC64;  
 Query Match 39.7%; Score 186.5; DB 1; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 1e-12;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
 OY 2 SGGVKYQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCMDW 61

```

Db      364 SGOIKVLYANKETNTTIRPWLKVNSSGSSIDLRSVTIRYWTVDGERAQAIS-DW 422
QY      62 AAMGCGNIRASFGSNVPATPTADTYLQ 88
Db      423 AQIGASNVTFKFKLSSSVSGADYLYE 449

RESULT 5
GUNA_CALSA
ID_GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RX SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: The linker region (also termed "hinge") may be a potential
CC site for proteolysis.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L32742; AAA91086.1; -
DR EMBL; M36063; AAA72860.1; -
DR EMBL; L01257; -; NOT_ANNOTATED_CDS.
DR PIR; T17120; T17120.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD_3; 2.

```

```

DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 39.7%; Score 186.5; DB 1; Length 1742;
Best Local Similarity 42.5%; Pred. No. 1.4e-12;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSPAGDNGQIKPGLQLVNTGSSVDLSITVTIRYWTVDGSGSTLVNCDW 61
Db 704 SGOIKVLYANKETNTTIRPWLKVNSSGSSIDLRSVTIRYWTVDGERAQAIS-DW 762
QY 62 AAMGCGNIRASFGSNVPATPTADTYLQ 88
Db 763 AQIGASNVTFKFKLSSSVSGADYLYE 789

RESULT 6
GUNA_CALSA
ID_GUNA_CALSA STANDARD; PRT; 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.L.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
CC F (family 10 of glycosyl hydrolases).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13602; CRA31936.1; -.  
DR PIR; S02711; S02711.  
DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD\_3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001000; Glyco\_hydro\_10.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00942; CEM\_3; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
DR PRINTS; PR00134; GLHYDRLASE10.  
DR ProDom; PD001947; CBD\_3; 1.  
DR SMART; SM00633; Glyco\_10; 1.  
DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;  
KW Multifunctional enzyme; Signal.  
FT SIGNAL 1 28  
FT CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.  
FT DOMAIN 376 416 THR/PRO-RICH, TANDEM REPEATS OF T-P.  
FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).  
FT DOMAIN 571 618 THR/PRO-RICH, TANDEM REPEATS OF T-P.  
FT ACT\_SITE 177 177 PROTON DONOR (POTENTIAL).  
FT ACT\_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 792 792 BY SIMILARITY.  
SQ SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;  
  
Query Match 38.8%; Score 182.5; DB 1; Length 1039;  
Best Local Similarity 41.9%; Pred. No. 2.1e-12;  
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
  
QY 3 GGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTDRDGSSTLVNCDWA 62  
DB 420 GQIKVLYANKETNTTIRPWLKVVNSGSSIDLSRTYRYWFTVDGERAQSAYS-DWA 478  
  
QY 63 AMGCGNIRASFGSNVPATPTADTYLQ 88  
DB 479 QIGASNVTFKFKLVSSVSGADYYLLE 504  
  
RESULT 7  
YCEA\_PAEIA STANDARD; PRT; 145 AA.  
AC Y29718;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein in celsa 5' region (Fragment).  
OS Paenibacillus lautus (Bacillus lautus).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Joergensen P.L., Diderichsen B.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22302.1; -.  
DR PIR; A41897; A41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Joergensen P.L., Diderichsen B.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22302.1; -.  
DR PIR; A41897; A41897.

DR HSSP; Q06851; INBC.  
DR InterPro; IPR001956; CBD\_3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR Pfam; PF00942; CEM\_3; 1.  
DR ProDom; PD001947; CBD\_3; 1.  
KW Hypothetical protein.  
FT NON\_TER 1 1  
SQ SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;  
  
Query Match 38.4%; Score 180.5; DB 1; Length 145;  
Best Local Similarity 43.9%; Pred. No. 3.4e-13;  
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;  
  
QY 7 VOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTDRDGSSTLVNCDWAAMGC 66  
DB 1 LQYRAADTNAADNQIKPSFNKNGTSAVDLSLKIRYFTKDGSAAVNGW-IDWAQLGG 59  
  
QY 67 GNIRASFGSNVPATPTADTYLQ 88  
DB 60 SNIQISFG--NHTGTNSDTYVE 79  
  
RESULT 8  
GUNA\_PAEIA STANDARD; PRT; 700 AA.  
ID GUNA\_PAEIA  
AC P29719;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase A) (EG-A).  
GN CELA.  
OS Paenibacillus lautus (Bacillus lautus).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.  
RX MEDLINE=92276330; PubMed=1592807;  
RA Hansen C.K., Diderichsen B., Joergensen P.L.;  
RT "celsa from Bacillus lautus PL236 encodes a novel cellulose-binding  
endo-beta-1,4-glucanase."  
RL J. Bacteriol. 174:3522-3531(1992).  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL  
CC HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M76588; AAA22303.1; -.  
DR PIR; B41897; B41897.

```

Query Match          35.2%; Score 165.5; DB 1; Length 879;
Best Local Similarity 33.0%; Pred.No 1,4e-10;
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

Qy 1 VSGGVKVVQYKNNDSAPGDNIQKPGLOLVNTGSSSVLSTTVRYWYTRDGGSSLTLYVNC D 60
db 736 IKGEVLYOANGAGATSNINPRFKINNGTKAINLSDVKIRYYTKYKGGASONPW-CD 794

```

```

795 WSSAGNSNVICGNFNLSPPAEGADICLE 822

DD
RESULT 10
CIPB_CLOTHM
ID_CIPB_CLOTHM STANDARD; PRT; 772 AA.
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein sl/sL)
DE (Cellulose interacting protein B) (Fragment).

```

```

CC -1- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY
CC THE CATALYTIC COMPONENTS OF THE CELLULOSE.
CC
CC -1- SIMILARITY: Contains at least 3 cohesin domains.
CC
CC -1- SIMILARITY: Contains 2 dockerin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X68233; CAA48312.1; -.
CC DR HSSP; X06851; INBC.
CC
CC DR InterPro; IPR001956; CBD_3.
CC DR InterPro; IPR008985; Cellul_bind.
CC DR InterPro; IPR002102; Cohesin.
CC DR InterPro; IPR002105; Dockerin_1.
CC DR InterPro; IPR002048; EF-hand_1.
CC DR pfam; PF00942; CBM_3; 1.
CC DR pfam; PF00963; Cohesin; 3.
CC DR pfam; PF00404; Dockerin_1; 2.
CC DR ProDom; PD001947; CBD_3; 1.
CC DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC DR PROSITE; PS00448; CL05 CELLULOSE_RPT; 2.
CC Cellulose degradation; Cell wall; Glycoprotein; Repeat.
CC
CC NON TER 1 1
CC DOMAIN <1 80 COHESIN 1.
CC FT 81 93 LINKER (PRO/THR-RICH).
CC FT

```

FT DOMAIN 94 240 COHESIN 2.  
 FT DOMAIN 241 272 LINKER (PRO/THR-RICH).  
 FT DOMAIN 273 439 CELLULOSE-BINDING.  
 FT DOMAIN 440 461 LINKER (PRO/THR-RICH).  
 FT DOMAIN 462 607 COHESIN 3.  
 FT DOMAIN 710 733 DOCKERIN 1.  
 FT DOMAIN 743 766 DOCKERIN 2.  
 SQ SEQUENCE 772 AA; 82491 MW; BFP06DE5E094FE10 CRC64;

Query Match 35.1%; Score 165; DB 1; Length 772;  
 Best Local Similarity 36.1%; Pred No. 1.4e-10;  
 Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPGDNQIKGLQLVNTGSSVDLSTVTVYVWFTDGGSSTLVYVNC 60  
 DB 277 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTRYVTVYVWFTDGGSS 335

QY 61 WAAM-----GGNTRASFGSVNPAIPTRADTYLQ 88  
 DB 336 HAAIIGNSYNGITSNVKGTFRKMSSTNNADTYLE 372

RESULT 11  
 CIPA\_CLOTH  
 ID CIPA\_CLOTH STANDARD; PRT; 1853 AA.  
 AC Q06851;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cellulosomal scaffolding protein A precursor (Cellulosomal  
 glycoprotein Sl/SL) (Cellulose integrating protein A) (Cohesin).  
 GN CIPA.  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1515;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.  
 RC STRAIN=ATCC 27405 / DSM 1237;  
 RX MEDLINE=93302508; PubMed=8316083;  
 RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,  
 RA Demail A.L.;  
 RT "Sequencing of a Clostridium thermocellum gene (cipA) encoding the  
 RT cellulosomal SL-protein reveals an unusual degree of internal  
 RT homology.";  
 RL Mol. Microbiol. 8:325-334(1993).  
 RN [2]  
 RP SEQUENCE OF 1820-1853 FROM N.A.  
 RX MEDLINE=9320931; PubMed=8458832;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulose to the cell surface.";  
 RL J. Bacteriol. 175:1891-1899(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.  
 RX MEDLINE=97238934; PubMed=9083107;  
 RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,  
 RA Frolow F.;  
 RT "A cohesin domain from Clostridium thermocellum: the crystal  
 RT structure provides new insights into cellulosome assembly.";  
 RL Structure 5:381-390(1997).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.  
 RX MEDLINE=97076134; PubMed=8918451;  
 RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,  
 RA Steitz T.A.;  
 RT "Crystal structure of a bacterial family-III cellulose-binding  
 RT domain: a general mechanism for attachment to cellulose.";  
 RL EMBO J. 15:5739-5751(1996).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.  
 RX MEDLINE=98022914; PubMed=9402065;

RA Tavares G.A., Beguin P., Alzari P.M.;  
 RT "The crystal structure of a type I cohesin domain at 1.7-A  
 resolution.";  
 RL J. Mol. Biol. 273:701-713(1997).  
 CC -!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT  
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
 CC CELLULOYTIC ENZYMES.  
 CC -!- SUBCELLULAR LOCATION: Cell surface.  
 CC -!- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY  
 CC THE CATALYTIC COMPONENTS OF THE CELLULOSE.  
 CC -!- SIMILARITY: Contains 9 cohesin domains.  
 CC -!- SIMILARITY: Contains 2 dockerin domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L08665; -; NOT ANNOTATED\_CDS.  
 DR EMBL; X67506; CAA47840.1; -;  
 DR FIR; S36859; S36859.  
 DR PDB; 1ANU; 23-JUL-97.  
 DR PDB; 1AOH; 08-JUL-98.  
 DR PDB; 1NBC; 26-SEP-97.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR008985; Cellul\_bind.  
 DR InterPro; IPR002102; Cohesin.  
 DR InterPro; IPR002105; Dockerin\_1.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00942; CBM\_3; 1.  
 DR Pfam; PF00963; Cohesin; 9.  
 DR Pfam; PF00404; Dockerin\_1; 2.  
 DR ProDom; PD001947; CBD\_3; 1.  
 DR PROSITE; PS00018; EF HAND; UNKNOWN 1.  
 DR PROSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
 KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;  
 KW 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1853 CELLULOSONAL SCAFFOLDING PROTEIN A.  
 FT DOMAIN 29 182 COHESIN 1.  
 FT DOMAIN 183 322 COHESIN 2.  
 FT DOMAIN 323 363 LINKER (PRO/THR-RICH).  
 FT DOMAIN 364 522 CELLULOSE-BINDING (BY SIMILARITY).  
 FT DOMAIN 523 559 LINKER (PRO/THR-RICH).  
 FT DOMAIN 560 704 COHESIN 3.  
 FT DOMAIN 724 866 COHESIN 4.  
 FT DOMAIN 889 1031 COHESIN 5.  
 FT DOMAIN 1054 1196 COHESIN 6.  
 FT DOMAIN 1219 1361 COHESIN 7.  
 FT DOMAIN 1384 1526 COHESIN 8.  
 FT DOMAIN 1548 1690 COHESIN 9.  
 FT DOMAIN 1791 1814 DOCKERIN 1.  
 FT DOMAIN 1824 1847 DOCKERIN 2.  
 FT CONFLICT 1615 1615 A -> AA (IN REF. 1).  
 FT STRAND 185 188  
 FT STRAND 190 191  
 FT TURN 195 196  
 FT TURN 198 206  
 FT TURN 210 211  
 FT STRAND 213 221  
 FT TURN 224 226  
 FT STRAND 227 234  
 FT TURN 236 237  
 FT TURN 243 246  
 FT STRAND 247 252  
 FT TURN 253 256  
 FT STRAND 257 263  
 FT TURN 265 266  
 FT TURN 270 271  
 FT STRAND 273 273

```
FT STRAND 277 286
FT TURN 292 304
FT STRAND 305 306
FT STRAND 309 309
FT STRAND 313 315
FT STRAND 317 319
FT STRAND 369 375
FT STRAND 381 382
FT STRAND 385 385
FT STRAND 387 393
FT STRAND 399 400
FT HELIX 401 403
FT STRAND 404 410
FT STRAND 418 428
FT TURN 430 431
FT STRAND 434 436
FT HELIX 438 440
FT STRAND 441 452
FT TURN 453 454
FT STRAND 455 463
FT STRAND 467 468
FT TURN 470 471
FT STRAND 473 482
FT TURN 483 484
FT STRAND 488 489
FT TURN 491 492
FT TURN 494 495
FT STRAND 498 498
FT STRAND 503 504
FT STRAND 509 512
FT TURN 513 514
FT STRAND 515 518
FT STRAND 1220 1224
FT STRAND 1226 1229
FT TURN 1231 1232
FT STRAND 1234 1242
FT TURN 1246 1247
FT STRAND 1249 1249
FT STRAND 1251 1257
FT TURN 1260 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT HELIX 1279 1282
FT STRAND 1283 1288
FT TURN 1289 1292
FT STRAND 1293 1299
FT TURN 1301 1302
FT TURN 1306 1307
FT STRAND 1309 1309
FT STRAND 1313 1322
FT TURN 1324 1325
FT STRAND 1329 1342
FT TURN 1344 1345
FT STRAND 1348 1348
FT STRAND 1351 1354
FT STRAND 1356 1360
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 35.1%; Score 165; DB 1; Length 1853;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLVYNC 60
Db 365 VSGNLKVFYFNPSDPTTNSINPQKFKVNTGSSAIDLSKLTLYRYVVDGQDKQTFW-CD 423
QY 61 WAAM-----GCNIRASFGSVNPTPTADTYLQ 88
Db 424 HAAIGSNGSYNGITSNVKGTFFVKSSSTNNADTYLE 460

RESULT 12
GUNW_ERWCA

ID AC Q59394; 277 286
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Atroseptica FCBR C18;
RA MEDLINE=9829944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multi-enzyme, which forms glucose.";
RL Biotechnology 14:71-76 (1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39788; AAC37033.1; -.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.8%; Score 149.5; DB 1; Length 444;
Best Local Similarity 38.6%; Pred. No. 3.8e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDG--GSSTLVYNC 59
Db 357 TGDVLYQYRNVNDNPNPSDDAIRMAVNIKNTGSTPKLSQLQVRYVYFHDGKFGANLTV--- 413
QY 60 DWAAAGCGNIRASFGSVNPTPTADTYL 87
Db 414 DWANVGPNNIVTSTGTPTAASDKNRYV 441

RESULT 13
GUNW_ERWCA
ID AC Q47056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
```

```
GN CELV.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: Endoglucanase with some exoglucanase activity.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
CC optimum about 42 degrees Celsius.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X76000; CAA53592.1; -.
DR PIR; S39962; S39962.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; Glyco_hydro_5.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 505 ENDOGLUCANASE V.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.8%; Score 149.5; DB 1; Length 505;
Best Local Similarity 38.6%; Pred. No. 4.4e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQYKNDSPAGNQIKPGLQVNTGSSVDLSTVTYVYFTRDG--GSSTLYVNC 59
Db 354 TGDVVLQYRNVNPNFSDDAIRMAFNKNTGSTPIKLSDLQVRYFHHDDGKPGANLFV--- 410

QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87
Db 411 DWANVGPNNIIVTGTTPAASDTKANRYV 438

RESULT 14
GUNW ERWCA
ID GUNW ERWCA STANDARD; PRT; 504 AA.
AC Q59195;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase VI)
DE (Cellulase VI).
```

```
GN CELV1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celV1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X79241; CAA55823.1; -.
DR PIR; S54744; S54744.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 504 ENDOGLUCANASE V1.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 31.4%; Score 147.5; DB 1; Length 504;
Best Local Similarity 37.5%; Pred. No. 7.4e-09;
Matches 33; Conservative 19; Mismatches 31; Indels 5; Gaps 2;

QY 2 SGGVKVQYKNDSPAGNQIKPGLQVNTGSSVDLSTVTYVYFTRDG--GSSTLYVNC 59
Db 354 TGDVVLQYRNVNPNFSDDAIRMAFNKNTGSTPIKLSDLQVRYFHHDDGKPGANLFV--- 410

QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87
Db 411 DWANVGPNNIIVTGTTPAASDTKANRYV 438

RESULT 15
GUX2 CLOSUR
ID GUX2 CLOSUR STANDARD; PRT; 914 AA.
AC P50900;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
DE (1,4-beta-cellobiohydrolase II) (Avicelase II).
GN CELY.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 10.385 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQYKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	257	54.7	741 16 Q82QF2	Q82qf2 streptomyc	
2	228.5	48.6	616 2 Q7X2N2	Q7x2n2 thermomon	
3	225.5	48.0	170 2 Q9RFK6	Q9rfx6 caldibacill	
4	216.5	45.1	930 2 Q9RFK5	Q9rfx5 caldibacill	
5	214.5	45.6	921 2 Q9L8L8	Q9l8l8 caldibacill	
6	209.5	44.6	1091 2 Q8KKF7	Q8kkf7 paenibacill	
7	204.5	43.5	997 2 Q9Z4I1	Q9z4i1 bacillus sp	
8	194.5	41.4	1751 2 Q9AQA4	Q9agg4 caldicellul	
9	192.5	41.0	1000 2 Q24820	Q24820 thermophili	
10	192.5	41.0	1770 2 Q9X3P5	Q9x3p5 caldicellul	
11	191.5	40.7	261 2 Q9AQG7	Q9agg7 caldicellul	
12	191.5	40.7	1426 2 Q9X3P6	Q9x3p6 caldicellul	
13	191	40.6	1711 2 P96311	P96311 anaerocellul	
14	190.5	40.5	996 2 Q9AQH0	Q9agh0 caldicellul	
15	190.5	40.5	1779 2 Q52374	Q52374 caldicellul	
16	190	40.4	499 2 Q93TJ6	Q93tj6 bacillus su	

17	190	40.4	508	2	Q93LD0	Q93ld0 bacillus su
18	187	39.8	499	2	O52731	O52731 bacillus sp
19	186	39.6	499	2	O45532	O45532 bacillus su
20	186	39.6	501	2	O83012	O83012 bacillus sp
21	180	38.3	486	2	O45430	O45430 bacillus sp
22	177	37.7	499	2	O8RPQ6	O8rpq6 bacillus am
23	172.5	36.7	1915	2	Q9RPL0	Q9rp0 acetivibrio
24	168	35.7	473	16	Q9RK75	Q9rk75 streptomyc
25	165.5	35.2	887	2	Q9L3J8	Q9l3j8 clostridium
26	157	33.4	542	2	Q7X3S6	Q7x3s6 bacillus li
27	150.5	32.0	2316	2	Q9FDJ9	Q9fdj9 bacteroides
28	130.5	27.8	1483	16	Q977Y4	Q977y4 clostridium
29	130.5	27.8	1546	2	O45996	O45996 clostridium
30	126.5	25.9	1162	2	O82830	O82830 clostridium
31	119.5	26.4	307	2	O46392	O46392 clostridium
32	117	24.9	1230	2	O59325	O59325 clostridium
33	107	22.8	221	2	O8VVI7	O8vvi7 clostridium
34	87	18.5	440	2	Q93LI9	Q93li9 bacillus sp
35	78	16.6	135	13	Q90542	Q90542 ginglymosto
36	76	16.2	121	13	O8JGA5	O8jga5 ginglymosto
37	76	16.2	121	13	O8JGA9	O8jga9 ginglymosto
38	75.5	16.1	618	4	O9Y3Z2	O9y3z2 homo sapien
39	75.5	16.1	660	2	Q9L3J2	Q9l3j2 clostridium
40	75.5	16.1	986	4	O94858	O94858 homo sapien
41	75.5	16.1	1300	4	Q8IZF2	Q8izf2 homo sapien
42	73.5	15.6	547	5	P91006	P91006 caenorhabdi
43	73.5	15.6	1428	16	O8YRU7	O8yru7 anabaena sp
44	73	15.5	511	5	O9GYG5	O9gyg5 caenorhabdi
45	72	15.3	499	12	Q993M3	Q993m3 autonomous

ALIGNMENTS

RESULT 1

ID	Q82QF2	PRELIMINARY;	PRT;	741 AA.
AC	O82QF2;			
DT	01-JUN-2003 (TREMELrel. 24, Created)			
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)			
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	Putative cellulose 1,4-beta-cellobiosidase.			
GN	GUXAI OR SAV557.			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RC	MEDLINE=21477403; PubMed=11572948;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis: deducing the ability of producing secondary			
RT	metabolites.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RC	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis.";			
RL	Nat. Biotechnol. 21:526-531(2003).			
DR	EMBL; AF005023; BAC6267.1; -			
DR	GO; GO:0004553; F:Hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001956; CBD 3.			
DR	InterPro; IPR008965; Cellul_bind.			
DR	InterPro; IPR003961; FN_III.			

```
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00441; fn3; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRLASE6.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00600; FN3; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Complete proteome.
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B4A7AF CRC64;

Query Match 54.7%; Score 257; DB 2; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.7e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSGSLTVVNCWDW 61
Db 591 SGLKLVYKNNDSATDNIAPGLRIVNTGSGSLDSKVYFYFTRDGSPTVNAWCYD 650

QY 62 AAMGCGNIRASFSGSVNPATPTADTYLQ 88
Db 651 AAVGCSNVSLKVPLTTPVPGADAYLE 677

RESULT 2
ID Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptoporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TW51;
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tfi cel5B, a new endoglucanase encoding
RT gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; --
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;

Query Match 48.6%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.8e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

QY 2 SGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSGSLTVVNCWDW 61
Db 471 TGALEYVYRNLSAADSQTAPGLRLVNTGSSVDLADVEIHYFTNEPG-GTLQFTCDW 529

QY 62 AAMGCGNIRASFSGSVNPATPTADTYL 87
Db 530 AQVGCANVASFTSL--SAPGADTSL 553

RESULT 3
ID Q9RFK6 PRELIMINARY; PRT; 170 AA.
AC Q9RFK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Caldibacillus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.
OX NCBI_TaxID=74586;
```

```

[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldibacillus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22273.1; --
DR HSP; Q06851; INBC.
GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 48.0%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.2e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 3 GGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSGSLTVVNCWDW 62
Db 20 GSLVQYRAADTNAGDNQLKPHFRIVNRGTSSVPLSELTIYWTVD-GDKPQVFNCDWA 78.

QY 63 AAMGCGNIRASFSGSVNPATPTADTYLQ 88
Db 79 QVGCNVRGSGFVKLSTORTGADYYIE 104

RESULT 4
ID Q9RFK5 PRELIMINARY; PRT; 930 AA.
AC Q9RFK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldibacillus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldibacillus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22274.1; --
DR HSP; Q06851; INBC.
GO: GO:001028; C:Viral capsid; IEA.
GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03087; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54DIA2CC CRC64;

Query Match 46.1%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 6.2e-15;
```

[illegible]

```
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR003961; FN III-like.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR000566; LipocIn_cytFABP.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 POTENTIAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.5%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.4e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGGVQVQKNDSPAGDQNIKPGQLQVNTGSSVDLSVTYVYWFTRDGGSTLVYNCDAW 61
Db 846 TCTLEVQVRSGGSGNSNAVTQFQNLKNTGTQAIIDLSTVKIRYPTKD-GTEELSPWCY 904

QY 62 AAMGCNIRASFGSVNPAFTADTYLQ 88
Db 905 AQGVSAVQGMFVAVNPARGKATDTVE 931

RESULT 8
Q9A0G4 PRELIMINARY; PRT; 1751 AA.
AC Q9A0G4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0003215; F:transporter activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR000566; LipocIn_cytFABP.
DR Pfam; PF00942; CBM 3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 3.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.4%; Score 194.5; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 3.4e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQVQKNDSPAGDQNIKPGQLQVNTGSSVDLSVTYVYWFTRDGGSTLVYNCDAW 63
Db 678 GVKVLYKNNETSASTGSRFWFKIVNGSSVDLSRVKIRYWTVDGDKPOSAY-CDWAQ 736

QY 64 MCGCNIRASFGSVNPAFTADTYLQ 88
Db 737 IGASNVTFNFVKLSSGVSGADYILE 761

RESULT 9
Q24820 PRELIMINARY; PRT; 1000 AA.
AC Q24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGKVQVQKNDSPAGDQNIKPGQLQVNTGSSVDLSVTYVYWFTRDGGSTLVYNCDAW 61
Db 372 SGQIKVLYANKETSTNTIRPWLKVNTGSSSIDLSRVIRYWTVDGDKAQSAYS-DW 430

QY 62 AAMGCNIRASFGSVNPAFTADTYLQ 88
Db 431 AQIGASNVTFKFVKLSSVSGADYILE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
ID Q9X3P5
QY 9X3P5
DB 9X3P5
SQ SEQUENCE 1770 AA; 192176 MW; 60178CBF3C00BE95 CRC64;
```

```
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE XYN.
GN XNA.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DR EMBL; AF078737; AAD30363.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR00584; CBD_IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_cenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02018; CBM_4_9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF03331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

OY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSSSTLVYNCDAWAA 63
DB 1104 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 1162

OY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 1163 IGASNVTFNFVKLSGSGVADYYLE 1187

RESULT 11
O9AQG7 PRELIMINARY; PRT; 261 AA.
AC O9AQG7;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Glycosyl hydrolase 6 (Fragment)
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

OY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSSSTLVYNCDAWAA 63
DB 1104 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 1162

OY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 1163 IGASNVTFNFVKLSGSGVADYYLE 1187

RESULT 11
O9AQG7 PRELIMINARY; PRT; 261 AA.
AC O9AQG7;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Glycosyl hydrolase 6 (Fragment)
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
```

```
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 261
FT NON_TER 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.7%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.3e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

OY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVRYWFTTRDGGSSSTLVYNCDAWAA 63
DB 118 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 176

OY 64 MCGNIRASFSGSVNPATPTADTYLQ 88
DB 177 IGASNVTFNFVKLSGSGVADYYLE 201

RESULT 12
O9X3P6 PRELIMINARY; PRT; 1426 AA.
AC O9X3P6;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DE CelB.
DR CELB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
```

```
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match
Best Local Similarity 40.7%; Score 191.5; DB 2; Length 1426;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGQLQVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDAWAA 63
Db 413 GLKVLTKNNETSASTGSIREFWFKLVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAAQ 471
QY 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

RESULT 13
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9493383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from the extreme thermophile Anaerocellum thermophilum with separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P26221; 1TF4.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM 3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD 3; 2.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
FT NON TER 1
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match
Best Local Similarity 40.6%; Score 191; DB 2; Length 1711;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-KVKVOYKNNDSAPGDNQIKPGQLQVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNC 59
Db 683 VAGGOIKVLKYANKETNSTTIRPWLKVVNTGSSVDLSRVKIRYWTVDGDKPQSAIS- 741
QY 60 DWAAAGCGNIRASFGSVNPTPTADTYLQ 88
Db 742 DWAAIGASNVTFKFKVLSGSGVADYYLE 770

RESULT 14
Q9AQH0 PRELIMINARY; PRT; 996 AA.
AC Q9AQH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P., Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6E2D2F3C614 CRC64;

Query Match
Best Local Similarity 40.5%; Score 190.5; DB 2; Length 996;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGQLQVNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDAWAA 63
Db 844 GLKVLTKNNETSASTGSIREFWFKLVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAAQ 902
QY 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGVADYYLE 927

RESULT 15
O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XVNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
```



DR GO; GO:0030246; F:carbohydrate binding; IEA.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR006584; CBD\_IV.  
 DR InterPro; IPR005084; CBM\_6.  
 DR InterPro; IPR003305; CBM\_cenc.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR008979; Gal\_bind\_like.  
 DR InterPro; IPR001000; Glyco\_hydro\_10.  
 DR Pfam; PF00942; CBM\_3; 3.  
 DR Pfam; PF02018; CBM\_4; 9; 2.  
 DR Pfam; PF03422; CBM\_6; 1.  
 DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
 DR PRINTS; PR00134; GLHYDRLASE10.  
 DR ProDom; PD001947; CBD\_3; 3.  
 DR SMART; SM00606; CBD\_IV; 1.  
 DR SMART; SM00633; Glyco\_10; 1.  
 DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
 KW Glycosidase; Hydrolase; Xylan degradation.  
 SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;  
 Query Match 40.5%; Score 190.5; DB 2; Length 1779;  
 Best Local Similarity 43.5%; Pred. No. 9.6e-12;  
 Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;  
 QY 4 GVKVOYKNDAPGDNOIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSTLVYNCWAA 63  
 Db 1113 GLKVLKKNETSASTGSRPWFKIVNGSSSSVDLSRVKIRIWTYVDGDKPSAV-CDWAQ 1171  
 QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88  
 Db 1172 IGASNVTFNFVKLTSGVSGADYILE 1196

Search completed: May 11, 2004, 12:09:28  
 Job time : 12.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 4.3675 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.psp.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.psp.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.psp.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.psp.\*

5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.psp.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.psp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.6	616	3	US-09-136-574A-47
2	194.5	41.4	1751	3	US-09-136-574A-44
3	191.5	40.7	1426	3	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PT-US95-13813-9
7	165	35.1	476	4	US-09-339-159B-4
8	165	35.1	493	3	US-09-198-956-10
9	165	35.1	493	3	US-09-198-955A-12
10	165	35.1	493	4	US-09-694-531-12
11	165	35.1	493	4	US-09-670-141-10
12	165	35.1	493	4	US-10-072-152-12
13	120.5	25.6	531	2	US-07-862-588B-7
14	114	24.3	162	1	US-08-048-164A-2
15	114	24.3	162	1	US-08-460-462-2
16	114	24.3	162	1	US-08-460-457-2
17	114	24.3	162	1	US-08-460-458-2
18	114	24.3	162	2	US-08-460-455-2
19	114	24.3	162	2	US-08-330-394A-2
20	114	24.3	163	3	US-09-006-636-7
21	114	24.3	163	3	US-09-006-632-7
22	114	24.3	163	4	US-09-325-274-7
23	113	24.0	382	3	US-09-277-716-22
24	113	24.0	382	4	US-09-609-161B-22
25	112	23.8	154	2	US-08-330-394A-29
26	112	23.8	156	2	US-08-330-394A-22
27	64	13.6	143	4	US-09-301-593-26

28	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appli
29	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
30	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
31	63.5	13.5	1785	4	US-09-341-587-3	Sequence 3, Appli
32	63	13.4	453	4	US-09-301-593-18	Sequence 18, Appl
33	63	13.4	472	4	US-09-301-593-43	Sequence 43, Appl
34	63	13.4	718	4	US-09-328-352-4640	Sequence 4640, Ap
35	63	13.4	1581	3	US-09-110-517-2	Sequence 2, Appli
36	62.5	13.3	288	4	US-09-423-439-38	Sequence 38, Appl
37	62.5	13.3	445	1	US-08-353-400-33	Sequence 33, Appl
38	62.5	13.3	641	4	US-09-687-538B-8	Sequence 8, Appli
39	62.5	13.3	673	4	US-09-423-439-32	Sequence 32, Appl
40	62.5	13.3	802	3	US-09-081-345-18	Sequence 18, Appl
41	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appli
42	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
43	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
44	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
45	61.5	13.1	307	3	US-08-370-476-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 616 amino acids

; TYPE: amino acid

; TOPOLOGY: single

; MOLFCULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:



```

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2

Query Match 36.6%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.le-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3

QY      1 VSGGVKVQVYKND-SAPGDNIQKPLQLVNTGSSVDLSVTVRYWFTRDGGSTLVNC 607
Db       | : | | | | : | | | | | : | | | | : | | | | : | | | | : | : | : |
        549 VNSDLVVQVKGRNATDNQIKPHFNQNGTSPVDLSSLTKLYIFTKD--SSAAMNGWI 607

QY      60 DWAAAGCGNIRASFGSVNPATPTADTYLQ 88
Db       ||| : | | : | | : | | : | | : | : | : | : | : | : | : | :
        608 DWAKLGGSNIQISFGNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-13813-9

Query Match 35.1%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 1.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSTLVYVNC 60
Db 5 VSGNLKVEFYNSPDDTINSINPQFKVTNTGSSAIDLSKLTLYRYTVVVGQDKQTFW-CD 63
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
Db 64 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 100

RESULT 7
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.1%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSTLVYVNC 60
Db 314 VSGNLKVEFYNSPDDTINSINPQFKVTNTGSSAIDLSKLTLYRYTVVVGQDKQTFW-CD 372
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
Db 373 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 409

RESULT 8
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSTLVYVNC 60
Db 331 VSGNLKVEFYNSPDDTINSINPQFKVTNTGSSAIDLSKLTLYRYTVVVGQDKQTFW-CD 389
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 426

RESULT 9
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
```

## RESULT 11

```

QY      62  AAMCGCNIRASFGSVNPATPTADTY 86
      | : | : | : | : | : | : | : |
Db      491 AQIGRTNVLLAF--ANFTGSNTDITY 513

RESULT 14
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-048-164A-2

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY      2  SGGVKVQYKNDGAPGDNQIKPGLQLVNTGSSVDLSTVTVRVYFTFDGGSSTLVYNCDW 61
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      3  TSSMSEFVYNSKSAQNTSITPIKINTGSDSLNLDNVKRVYYTSDTGQGTFFW-CDH 61

QY      62  AAMCGCN-----IRASF--GSVNPATPTADTYLQ 88
      | | | | | | | | | : | : | : | : | : | : | : | : |
Db      62  AGALLGNSVYDNTSKVTANFVKETASP-TSTYDTIVE 97

RESULT 15
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Dol, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21

```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

```

```

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. NO. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY      2 SGGVKVQYKNDSPGDNQIKGLQLVNTGSSVDLSTVTYRYWETRDGSSSTLVYNCDW 61
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 TSSMSVEFYFNKSAQTSITPIIKITNTSDNLNDVKRYYYTSDGTQGTFW-CDH 61
QY      62 AAMGCGN-----IRASF--GSVNPATPTADTYLQ 88
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 AGALLGNSYVDNTSKVTANFVKETASP-TSYDTYVE 97

```

Search completed: May 11, 2004, 12:11:21  
Job time : 5.3675 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 10.385 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US05\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	12	US-09-917-376-5
2	469	99.8	88	14	US-10-155-400-5
3	469	99.8	89	12	US-09-917-376-4
4	469	99.8	89	14	US-10-155-400-4
5	469	99.8	154	10	US-09-917-378-4
6	469	99.8	762	10	US-09-917-378-1
7	466	99.1	150	10	US-09-917-384-5
8	466	99.1	150	10	US-09-917-383-5
9	466	99.1	1043	10	US-09-917-384-6
10	466	99.1	1043	10	US-09-917-383-6
11	466	99.1	1228	10	US-09-917-384-1
12	466	99.1	1228	10	US-09-917-383-1
13	462	98.3	957	12	US-09-917-376-1
14	462	98.3	957	14	US-10-155-400-1
15	257	54.7	741	14	US-10-156-761-8100

16	190	40.4	508	15	US-10-369-493-23151	Sequence 23151, A
17	180	38.3	1621	14	US-10-185-990-10	Sequence 10, Appl
18	165	35.1	476	12	US-10-372-054-4	Sequence 4, Appl
19	165	35.1	493	12	US-10-655-433-12	Sequence 12, Appl
20	165	35.1	493	13	US-10-072-152-12	Sequence 29, Appl
21	165	35.1	599	10	US-09-955-555A-29	Sequence 2, Appl
22	155.5	33.1	1352	10	US-09-784-554B-2	Sequence 4, Appl
23	150.5	32.0	1350	10	US-09-784-554B-4	Sequence 4, Appl
24	130.5	27.8	1483	12	US-10-282-122A-51483	Sequence 51483, A
25	114	24.3	256	14	US-10-261-446-6	Sequence 22, Appl
26	113	24.0	382	10	US-09-808-898-22	Sequence 2, Appl
27	105.5	22.4	163	12	US-10-460-524-2	Sequence 2, Appl
28	78.5	16.7	1049	12	US-10-282-122A-49900	Sequence 49900, A
29	75.5	16.1	618	12	US-10-211-462-223	Sequence 53, Appl
30	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appl
31	75.5	16.1	986	12	US-10-312-312-53	Sequence 101, App
32	75.5	16.1	986	14	US-10-120-604-101	Sequence 406, App
33	75.5	16.1	986	14	US-10-225-567A-406	Sequence 3, Appl
34	75.5	16.1	986	16	US-10-398-458-3	Sequence 24, Appl
35	75.5	16.1	1131	16	US-10-398-458-2	Sequence 20, Appl
36	75.5	16.1	1325	9	US-09-747-835A-24	Sequence 61, Appl
37	75.5	16.1	1325	12	US-10-312-312-24	Sequence 20, Appl
38	75.5	16.1	1346	9	US-09-747-835A-20	Sequence 61, Appl
39	75.5	16.1	1346	9	US-09-747-835A-61	Sequence 25, Appl
40	75.5	16.1	1346	12	US-10-312-312-20	Sequence 25, Appl
41	75.5	16.1	1346	12	US-10-312-312-61	Sequence 196885, App
42	75.5	16.1	1371	9	US-09-747-835A-25	Sequence 377, App
43	75.5	16.1	1371	12	US-10-312-312-25	
44	75	16.0	161	12	US-10-424-599-196885	
45	72.5	15.4	69	9	US-09-764-860-377	

ALIGNMENTS

RESULT 1

US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

Query Match	99.8%	Score 469;	DB 12;	Length 88;
Best Local Similarity	100.0%	Pred. No. 1.5e-48;		
Matches	88;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSLTLYNCD	60	
Db	1	VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVWFTRDGSSLTLYNCD	60	
QY	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88	
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88	
RESULT 2				

```
US-10-155-400-5
; Sequence 5, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match          99.8%; Score 469; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60
DB 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match          99.8%; Score 469; DB 12; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60
DB 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

US-10-155-400-4
; Sequence 4, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-4

Query Match          99.8%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60
DB 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSLTVTVYWFTRDGGSSTLVYVNC 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match          99.8%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 6.8e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
Db 477 VSGGLKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 536

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 537 WAAMGCGNIRASFGSVNPATPTADTYLQ 564

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 6.8e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
Db 477 VSGGLKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 536

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 537 WAAMGCGNIRASFGSVNPATPTADTYLQ 564

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
Db 584 VSGGLKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 643

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 60
Db 584 VSGGLKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYNCD 643

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
```

; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 98.3%; Score 462; DB 12; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVYXCD 60  
DB 869 VSGGVKQVQKXNDSPAGDNQIKPGLQVYNTGSSVDLSTVTYRYWFTRDGGSSTLVYXCD 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14  
US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 98.3%; Score 462; DB 14; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVYXCD 60  
DB 869 VSGGVKQVQKXNDSPAGDNQIKPGLQVYNTGSSVDLSTVTYRYWFTRDGGSSTLVYXCD 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15  
US-10-156-761-8100  
; Sequence 8100, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8100  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100

Query Match 54.7%; Score 257; DB 14; Length 741;  
Best Local Similarity 56.3%; Pred. No. 6e-22;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
QY 2 SGGVKVQYKXNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVYXCDW 61  
DB 591 SGGKLVLYKXNDSSATDNAIRPGLRIVNTGSGSLDLSKVTRYVYFTRDGSGSPTVNAWCDY 650  
QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 651 AAVGCSNVLKVVPLTTPVPGADAYLE 677

Search completed: May 11, 2004, 12:25:48  
Job time : 11.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 1279.29 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQYKNDSPAGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frames\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_7742@runat\_11052004\_114535\_28241 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.to.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

RESULT 1

ALIGNMENTS

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	469	99.8	2289	6	AX700050	Sequence
2	466	99.1	3687	6	AX700036	Sequence
3	462	98.3	2869	6	AX700058	Sequence
4	436	92.8	3365	6	AX700025	Sequence
5	257	54.7	299175	1	AP005023	Streptomy
6	228.5	48.6	1357	1	AY298814	Thermobif
7	225.5	48.0	4567	1	AF163837	Caldibaci
8	214.5	45.6	3237	1	AF200304	Caldibaci
9	209.5	44.6	3509	1	PAE488933	Paenibaci
10	204.5	43.5	4161	1	BSP133614	Bacillus
11	195.5	41.6	2029	6	E35142	Truncated c
12	195.5	41.6	2029	6	E35143	Truncated c
13	194.5	41.4	6005	1	AF078038S5	Caldicell
14	194.5	41.4	6416	6	E35100	Truncated c
15	192.5	41.0	4743	1	AB008029	Thermophi
16	192.5	41.0	11707	1	AF078737	Caldicell
17	192.5	41.0	11707	6	E35099	Truncated c
18	191.5	40.7	787	1	AF078038S3	Caldicell
19	191	40.6	5513	1	ATZ86105	A.thermophi
20	190.5	40.5	3262	1	AF078038S1	Caldicell
21	190.5	40.5	5437	1	CASR69XN2	Caldicell
22	190	40.4	1523	1	AF355629	Bacillus
23	190	40.4	1800	1	BSEGLSG	Bacillus su
24	190	40.4	1920	1	BACGLUB	B. subtilis
25	190	40.4	2314	1	BS14GLUC	B. subtilis
26	190	40.4	2435	1	BSBGLUC2	Bacillus su
27	190	40.4	2589	1	AY044252	Bacillus
28	190	40.4	26170	1	BC170DEGR	B. subtilis
29	190	40.4	207829	1	BSUB0010	Bacillus su
30	188	40.0	1928	1	BACMCASE	D01057 B. subtilis
31	187	39.8	1593	1	AY183475	Bacillus
32	187	39.8	2084	1	AF045482	Bacillus
33	186.5	39.7	4977	1	CDCMANA	Caldocellum
34	186.5	39.7	5439	1	CDCELA	Caldocellum
35	186	39.6	2175	1	BACCELD	B. subtilis
36	186	39.6	2175	6	E05425	DNA sequenc
37	186	39.6	2587	1	AB016164	Bacillus
38	182.5	38.8	2977	6	A28170	B. lautus st
39	182.5	38.8	4241	1	CSCELB	Caldocellum
40	182.5	38.8	5284	1	CSU16308	Caldocellum
41	180.5	38.4	2831	1	BACCELA	Bacillus la
42	180	38.3	1553	1	BSU27084	Bacillus sp
43	177	37.7	2152	1	AF363635	Bacillus
44	172.5	36.7	6555	1	AF155197	Acetivibr
45	168	35.7	291000	1	SC0939105	Streptomy

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES



```

AX700050
LOCUS AX700050 2289 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012110.
ACCESSION AX700050
VERSION AX700050.1 GI:29536020
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS CELLULOLYTICUS
Patent: WO 03012109-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..2289
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,218-48 Length: 2289
Score: 469.00 Matches: 86
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700050 (1-2289)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACCGAG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1423 ATCAACCGCGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTTCGACGGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGGCGGTACTGGTTACCCGGATGTGGTCTGCACACTGGTGTAACCTGTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGGCGCGGATGGGTGTGGATATCCGGCTCGTTCGGTTCGGTGAACCGCGGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS CELLULOLYTICUS
Patent: WO 03012109-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of Guxa"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 9,538-48 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700036 (1-3687)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACCGAG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1810 ATCAACCGCGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACCGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGCGGTACTGGTTACCCGGATGTGGTCTGCACACTGGTGTAACCTGTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGGCGCGGATGGGTGTGGATATCCGGCTCGTTCGGTTCGGTGAACCGCGGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013

RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS CELLULOLYTICUS
Patent: WO 03012090-A 2 13-FEB-2003;
JOURNAL Midwest Research Institute (US)
FEATURES
source
1..2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,198-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservatives: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700058 (1-2869)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGTGATAACCGAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGCGGTTCGACGTGGTGAATACCGGGTCTGTCGGTGGATTTCGACCGGTG 2724

```

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 2725 ACCTGGCGGACTACTGTTCAACCGGGATGGTGGCTCGTCGACACTGGTGTACAACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 2785 TGGGGCGGCGATCGGGTGTGGGAATATCCGGCCCTCGTTCCGCTCGGTGAACCCGGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 2845 CCGACGGCGACACTACTCTGCAG 2868

RESULT 4  
AX700025 3365 bp DNA linear PAT 03-APR-2003  
LOCUS Sequence 2 from Patent WO03012095.  
DEFINITION AX700025  
ACCESSION AX700025  
VERSION AX700025.1 GI:29536018  
KEYWORDS  
SOURCE  
ORGANISM  
Acidothermus cellulolyticus  
Acidothermus cellulolyticus  
Bacteria: Actinobacteria: Actinobacteridae; Actinomycetales;  
Frankineae; Acidothermaceae; Acidothermus.

REFERENCE  
1 Adney,W.S., Ding,S.Y., Vinzant,T.B., Himmel,M.E., Decker,S.R. and Lantz McCarter,S.  
Thermal tolerant exoglucanase from Acidothermus cellulolyticus  
Patent: WO 03012095-A 2 13-FEB-2003;  
JOURNAL Midwest Research Institute (US)  
FEATURES  
source  
1. 3365  
/organism="Acidothermus cellulolyticus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:28049"

ORIGIN  
Alignment Scores:  
Pred. No.: 4.72e-44 Length: 3365  
Score: 436.00 Matches: 81  
Percent Similarity: 97.62% Conservative: 1  
Best Local Similarity: 96.43% Mismatches: 2  
Query Match: 92.77% Indels: 0  
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700025 (1-3365)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
Db 112 CTCAAAGCGCAGTATAAGAACATGATTCGGCGCGGAGTGACACAGATCAACCGGGT 171

QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
Db 172 CTCACAGTGGTGAATACCGGTCTGTCGGTGGATTTGTTCGACGGTACCGTGGCGTAC 231

QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyrAlaAlaMet 64  
Db 232 TGGTTACCCCGGATGGTGGGTCTGTCGACACTGGTGTACACTGTGATCGCGCGGATG 291

QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
Db 292 GGGTGTGGGAATATCCGGCCCTCGTTCCGCTCGGTGAACCCGGCGACCGCGCGGAC 351

QY 85 ThrTyrLeuGln 88  
Db 352 ACCTACTGTGCAG 363

RESULT 5  
AP005023 299175 bp DNA linear BCT 10-MAY-2003  
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 3/30.  
DEFINITION AP005023  
ACCESSION AP005023 BA000030  
VERSION AP005023.1 GI:29604083

KEYWORDS  
SOURCE  
ORGANISM  
Streptomyces avermitilis MA-4680  
Streptomyces avermitilis MA-4680  
Bacteria: Actinobacteria: Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE  
1 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.  
Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
11572948

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
2  
AUTHORS  
Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.  
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
3 (bases 1 to 299175)  
AUTHORS  
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kishida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nite.go.jp, URL: http://www.bio.nite.go.jp/, Tel.81-3-3481-1933, Fax:81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi Osonoe(\*4), Norihiro Kishida(\*4), Hisashi Kikuchi(\*4), Tadayoshi Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7) and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermitilis.ls.kitasato-u.ac.jp.  
Location/Qualifiers  
1. 299175  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"  
complement(155..550)  
/note="SAV445"  
Tn4561, truncated"  
/codon\_start=1  
/transl\_table=11  
/product="putative transposase"  
/protein\_id="BAC68155.1"  
/db\_xref="GI:29604084"  
/translation="MAGPPRSATRSRTTADREDSAHAPADPEPGYRRQIKVQANLQ DEHVALRRAGLQYRQDGMEDQIGALGLVNLVLFNTRYMDALTQLRADGFEVR DEDVALRSPFMRHHNVLGRYSFQLPDL"  
complement(474..740)  
/note="IS630 family, truncated"

FEATURES  
source  
CDS  
CDS



DB: 1 2 Gaps:

US-09-917-376-4 (1-89) x AY298814 (1-1957)

Qy 2 SerGlyClyVallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnile 21  
 Db 1517 ACGGGTGGCTCGAGGTCTATACCCGCAACACAGTCTTGGCGCGACGACGACCGATC 1576  
 Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
 Db 1577 GCGCGGGGCTGGTCTGGTCAACACCGAAGCAGCACGGTCGCTGCTGCGTGGAA 1636  
 Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 1637 ATCCACTACTACTTCAACACGAGCCGGC---GGTACCTCCAGTTCACCTGCGACTGG 1693  
 Qy 62 AlaAlaMetClyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
 Db 1694 GCTCAAGTGGGTGCGCCCAACGTCACCGGTCTTCCAGTCGCTG-----TCGGCTCCG 1747  
 Qy 82 ThrAlaAspThrTyrLeu 87  
 Db 1748 GCGCGCGACACTCCCTG 1765

RESULT 7

AF163837

LOCUS

DEFINITION

AP163837 4567 bp DNA linear BCT 08-FEB-2000

Caldibacillus cellulovorans multidomain beta-1,4-mannanase

precursor (manA) gene, complete cds; and unknown genes.

ACCESSION

AF163837

VERSION

AF163837.1 GI:6651325

KEYWORDS

SOURCE

ORGANISM

Caldibacillus cellulovorans

Caldibacillus cellulovorans

Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;

Caldibacillus.

REFERENCE

1 (bases 1 to 4567)

Sunna,A., Gibbs,M.D., Chin,C.W., Nelson,P.J. and Bergquist,P.L.

A gene encoding a novel multidomain beta-1,4-mannanase from

Caldibacillus cellulovorans and action of the recombinant enzyme on

kraft pulp

Appl. Environ. Microbiol. 66 (2), 664-670 (2000)

20120520

PUBMED

10653733

REFERENCE

2 (bases 1 to 4567)

Sunna,A., Gibbs,M.D. and Bergquist,P.L.

Direct Submission

Submitted (29-JUN-1999) Biological Sciences, Macquarie University,

North Ryde, New South Wales 2109, Australia

FEATURES

source

1..4567

/organism="Caldibacillus cellulovorans"

/mol\_type="genomic DNA"

/db\_xref="taxon:74586"

<1..513

/note="orf1"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAF22273.1"

/db\_xref="GI:6651326"

/translation="TPTPTPTPSAAPTPTPTPSAGSLVVOYRAADTNAGDNOLKPHPR1

VNRGTSVPLSELTIRYWTVDGDKPQVNCDAWGQVGSNVRGSFVKLSTGRGTADY

IEITPSGAGSLAAGASSGDIQIRINKDWTYNENDYSYDPTKTSFADNRVTLIR

NGQLIWGVRP"

misc\_feature

1..60

/note="Region: proline-threonine linker"

/evidence=not\_experimental

misc\_feature

61..510

/note="Region: cellulose-binding domain type IIb"

/evidence=not\_experimental

620..3412

/gene="manA"

CDS	620..3412 /gene="mana" /codon_start=1 /transl_table=11 /product="multidomain beta-1,4-mannanase precursor" /protein_id="AAF22274.1" /db_xref="GI:6651327"	DB:	1	Gaps:	1
	/translation="MNRRLIARLSGLMLAWLIAVLAVVPKPEAEAHGGMVFPATRT YACGYVDGKHGNGDNLNINPACDLALISGNYQFWMFGLNISNAGGRHREIIPDGK LCGPTASPDGNQARTDWMTRLRLOPGATITVRVNAWAPFGTWLYLVTRDGDWTPQPL KWSLEPFPFVQNPPTNSGPDGAESYQVOLPNKGRHIIYMIWORSDFEAFYN CSDVYEGSGPIAERGEDRGSTWITPPSGTPTPTPTPTPTPTPTPTPTPTPTPTPT TPTSTPTPSAGSTLAREVRGDTSATDNQMPQLAIVNTGSOAVLELTKRYWYTKN STQAEYFCDDWAQICGSNIRAQFVLSQFVSGADSYIELSFSGSIPAGNTEIQNR IHFTNMNNTEDMSYNAQTWGPSTRIILYRNGVLVWGTPEPGSGSPPTPTPTPT PT GANAVRIVLSNGCRWTKIPASEVADIISOARTGLAVRVLEVHDTTGVGEDAAACSMT TAVNYIELKVLAGEQNFVIVNIGNEPYGNVYONWYDTEFNVAOLRNAGINNTIM VDAPNWGDWSTWRDNPATPTFNADPQRLNLFVSIHNGYVDTAEEVQSYIESFVNRGL PLVIGFEGHMSDFDNEQAIQVYAKQINIGLFGWSGNGGGVEYLDVMTNFNANS TAWGTWPTNAIGTSTPT KPHFRIVNRGTSGSVPLSELTRYWYTVGDKPQVFNCDWAQVGCNLRGSGFVKLSTGR TGADYIIRITFTSGAGSLAPGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADWN RVTLRYNQLVWGVEP"	US-09-917-376-4 (1-89) x AF163837 (1-4567)			
sig_peptide	620..718 /gene="mana" /evidence=not_experimental	Qy	3	GlyGlyValIysValGlnTyrIysAsnAsnSerValAlaProGlyAspAsnGlnIleIys	22
mat_peptide	719..3409 /gene="mana" /product="multidomain beta-1,4-mannanase" /gene="mana" /note="Region: proline-threonine linker"	Db	58	GGCAGCTCGTGTGAGTATCGCGCGGCGACACCAACGCGGCGACACCACTGAAG	117
misc_feature	1349..1480 /product="multidomain beta-1,4-mannanase" /gene="mana" /note="Region: proline-threonine linker"	Qy	23	ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal	42
misc_feature	1481..1927 /gene="mana" /note="Region: cellulose-binding domain type IIib"	Db	118	CCGATTTTTCGATTGTGAACCGCGGACGTCGAGCGTGCCTGTCGGAGCTCACGATT	177
misc_feature	1928..2032 /gene="mana" /note="Region: proline-threonine linker"	Qy	43	ArgTyrTirPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirPala	62
misc_feature	2033..2884 /gene="mana" /note="Region: proline-threonine linker"	Db	178	CGGTACTGTGATCCGGTGGAC---GGGCAACGCCAGGTGTCACTGCGACTGGCGG	234
misc_feature	2885..2959 /gene="mana" /note="Region: cellulose-binding domain type IIib"	Qy	63	AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr	82
misc_feature	2960..3409 /gene="mana" /note="Region: cellulose-binding domain type IIib"	Db	235	CAGTGGGTGTTTCGAACTGCGCGGACGTTCTCGAAGCTTTCGACGGCGCGACGGGG	294
CDS	3894..4400 /note="Orf3" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAF22275.1" /db_xref="GI:6651328"	Qy	83	AlaAspThrTyrIleuGln	88
	/translation="MGASGVFSRTAQNGLKARQKRCRCVGRNLLGHLRFVDDDEF RFFAGARLRGRFSRQNLGEAQNFEQSTVTTFVQDDKRRIRRNIGICGLVRFQASII GPKRLIEAEQQAERQASDQPOLKNVILHFIDQHPYGVVERMDPFHKEIQKS RSILTRYI"	Db	295	GCGGACTACTATATCGAG	312
Alignment Scores:	1.3e-17 Score: 225.50 Percent Similarity: 67.44% Best Local Similarity: 47.67% Query Match: 47.98%	RESULT 8			
		AF200304			
		LOCUS			
		DEFINITION			
		ACCESSION			
		VERSION			
		KEYWORDS			
		SOURCE			
		ORGANISM			
		REFERENCE			
		AUTHORS			
		TITLE			
		JOURNAL			
		MEDLINE			
		PUBMED			
		REFERENCE			
		AUTHORS			
		TITLE			
		JOURNAL			
		FEATURES			
		source			
		CDS			
		gene			
		CDS			

```
/protein_id="AAF61649.1"
/db_xref="GI:7385020"
/translation="MGTRAGEWIKKMLLVLAIGLLIPYPHYVASENVLLIQSDFE
DGTQGVGGGCVETLTVTSAAAYSGAYLSVSGRTETWHGPTLIDTYSYQVGTQYF
SAWKLQSGGNTSISMTMOTMODTVYEOIYFDTALSGNWIOLKQYKLYEPAVNL
QVYEPADHATQSYVIDDVRLEQ.PDLKPTVEENIPSLKQVFAGRFPIGTAFENFELL
DEQKLLIKHFNSTPGNVLKWDSTPQEGVFNFTESDKAVAFVQNGMKIRGHTLI
WHNQTPMNVFYDSNGNLVSKEVLYORMERHIKTVVSRKGIYIYADVNVNEVIDPQPD
GLRSLWQIAGEYIEKAFQAEADPNALLFINDVNTHESGSQALYNIQVRLKSK
GIPVHGVOHONITWPSIEISFNKVLGVOBEITELDSIYNNSSQKDYTLPL
SDLAQOQATRVQLPEFLRSSLIONVTFGKDDANTWLRKFPVVRNDWLLRDEQL
KAKPAYNAVGVSTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
RVYRVDSSAIDNQMKQLRIVTVGSQAVPLTELKRYIKSTQAEQYPCDWAQI
CSNIRAQFVSLAQSADSGSYIELSTFGSPAGGNTGEIQNRHFTNMWNYNETDD
WSYNGTQTWGPSRITLIYRNGVLVWGTEPGGSGSTPTPTPTPTPTPTPTPTPTPT
STPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
GTTSPVLSSELTIRWYTVDGKQVFNCDMAWVGSNLGRSLVLTGRTGADYVLEI
TFTSGAGSLAPGANSGLDIQARINKNDWNTYNEANDYSYDPTKTSFADNRVTLRYNGQ
LVWGVEP"
sig_peptide 384..482
mat_peptide 483..3146
misc_feature /gene="xynA"
misc_feature /product="beta-1,4-xylanase XynA"
misc_feature /score=.965
misc_feature /gene="xynA"
misc_feature /note="Region: thermostabilizing domain"
misc_feature 966..1946
misc_feature /gene="xynA"
misc_feature /note="Region: beta-1,4-xylanase catalytic domain"
misc_feature 1947..2075
misc_feature /gene="xynA"
misc_feature /note="Region: proline-threonine linker"
misc_feature 2076..2519
misc_feature /gene="xynA"
misc_feature /note="Region: cellulose-binding domain typeIIb"
misc_feature 2520..2696
misc_feature /gene="xynA"
misc_feature /note="Region: proline-threonine linker"
misc_feature 2697..3146
misc_feature /gene="xynA"
misc_feature /note="Region: cellulose-binding domain typeIIb"
ORIGIN
Alignment Scores:
Pred. No.: 1.99e-16 Length: 3237
Score: 214.50 Matches: 40
Percent Similarity: 65.12% Conservative: 16
Best Local Similarity: 46.51% Mismatches: 29
Query Match: 45.64% Indels: 1
DB: 1 Gaps: 1
US-09-917-376-4 (1-89) x AF200304 (1-3237)
QY 3 GlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLys 22
Db 2694 GCAACATTTTCGATGTCAGTACCGCGGCGGACACAAACCGCGGACAGCGTGAAG 2753
QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 2754 CGCAGATTTTCGATGTCAGTACCGCGGCGGACGACGCGTACCGTTCGAGGCTTACGATT 2813
QY 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62
Db 2814 CGGTACTGGTACACGGTGGAC---GCCGACACAGCGGAGGTGTTCAACTGTGACTGGCGG 2870
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 2871 TGGGTTCGATGTCAGTACCGCGGCGGAGTCTGGTGAAGTTGACGACGCGGCGGCGG 2930
QY 83 AlaAspThrTyrLeuGln 88
Db 2931 GCGGACTACTACTCTTGAG 2948
```

```
RESULT 9
PAE488933 3509 bp DNA linear BCT 15-JUN-2002
LOCUS Paenibacillus sp. BP-23 cel48C gene for cellulose
DEFINITION 1,4-beta-cellobiosidase.
ACCESSION AJ488933
VERSION AJ488933.1 GI:21449823
KEYWORDS cel48C gene; cellulose 1,4-beta-cellobiosidase.
SOURCE Paenibacillus sp. BP-23
ORGANISM Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
REFERENCE 1
AUTHORS Sanchez,M.M., Pastor,F.I.J. and Diaz,P.
TITLE Paenibacillus sp. BP-23 family 48 Cellulase. Cloning and
perforamce on cellulosic substrates
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3509)
AUTHORS Diaz,P.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Diaz P., Microbiology, University of
Barcelona, Av. Diagonal, 645, 08028-Barcelona, SPAIN
FEATURES
source
1..3509
/organism="Paenibacillus sp. BP-23"
/mol_type="genomic DNA"
/strain="BP-23"
/db_xref="taxon:198119"
/country="Spain:Catalonia, Ebro river delta"
20..3472
/gene="cel48C"
20..25
/gene="cel48C"
20..25
/gene="cel48C"
39..45
/gene="cel48C"
62..3337
/gene="cel48C"
/EC_number="3.2.1.91"
/function="cellulose hydrolysis"
/codon_start=1
/transl_table=11
/product="cellulose 1,4-beta-cellobiosidase"
/protein_id="CAD32945.1"
/db_xref="GI:21449824"
/db_xref="GOA:Q8KKF7"
/db_xref="SPTREMBL:Q8KKF7"
/translation="MLKTAAKSLTAMLAGTVMLTGYTGLTGSQTAAHSDQAIEIOA
DGINEARFLQVLAQLKDPANGYFSAERGIPVHSETLILSEAPDYGHMSTSEAYSVWLWL
ETMYGHYTGDSKLEAANDMEKYIIPVNEGDNESQPTMSKYNPSPATYAAEKPP
DQYPSNLNGQYAAKQDPLDAELKATYNNQTYLHMWLIIDVNNWYGLNLLNPSTHTY
VNTFQREGESYWEAIPHPSQDDKSGKAGEGMSLFTKESQAPSAQWRYNTATDADA
RAVQVLYWAKENYNNSEYLDKAKKMGDFLYGMYDKYFKQVGSAXNGTPTGTDKDS
NNYLMAWTTSWGGGLQGGDWAIRICASHTHOAYONPVAAYALSDPAGGLIPDSPTAK
ADWNTATKEQLFEYTWLOSHEGAVGGATNSIGSYAAYPAGVSTFVDMAYQAPVVR
RVDSTWTFGQFAPLERVAEMTYIILAESGLISENFQMAKKVILTKVDWTKDYFVGE
REVTDAQGYLNAAGRIILGGTNAQVATTAPAGPEFWIPGGQEQGQDDKWNFGSTYTN
NNFNHAIKTDPQDQTVGLGYIKALTFFAAGTEAENGTLISAKQEKDQAQSLDITAM
YIGSDIAPAIKQDPAWAYLDNKYKTSYNTPTTKQWENGAPTFTYHRFWSQVDNATAYG
EFDRLLIGSSSTFGVQVPSAPAGVTAAGDKQVVLWNAAAGATSYSVKATVNGGPTT
SVATQVLTGSTFTDITGLTNGTAYIYVITAVNSAGESAPSAQVATPQAGVTVPGVTLUS
GNAGDAQVLTNNWTASTGAATYTVQSTGTGAVANLATGLTALTYYTDTTAVNGTAYNR
VATWASGOTNSNVVLTTPMAAPISTGTLEQVYRNGGSSASGNATIPQFNLKNTGTTA
IDLKSVKRYYYTKDSAADSWFCVDAQLSANVOGSFVAVNPAGKTADTYLBLELSPA
GAGSLAAGATGIIQTRFSKNWNTNFDQSDNYSFNATQTAFAGATKVTVAQDGAKVWG
MEP"
sig_peptide 62..166
mat_peptide /gene="cel48C"
167..3334
/gene="cel48C"
/product="cellulose 1,4-beta-cellobiosidase"
```

```

/EC_number="3.2.1.91"
3456..3472
/gene="celB"

terminator
ORIGIN
Alignment Scores:
Pred. No.: 9,26e-16 Length: 3509
Score: 209.50 Matches: 40
Percent Similarity: 70.11% Conservative: 21
Best Local Similarity: 45.98% Mismatches: 25
Query Match: 44.57% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x PAE488933 (1-3509)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 2879 ACGGGAACGCTGGAAGTGCAGTATCGCAATGAGGGTCCAGGTCCTTCAGTATGCAATT 2938
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 2939 ACGCCGCAATTAAATCTGAAAATAACAGGTACACGGCAATTGATCTGAGTAAGGTGAAG 2998
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 2999 GTTCGTTATTATTTCCAGGAAGAC---ACCGCAGCGGACATGAGCTTCTGGTGTGACTAT 3055
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 3056 GCACAGCTGGCAGTCCCAATGTGAGGCGACCTTGTGCTGTAAATCCGCGAAGGCG 3115
QY 82 ThrAlaAspThrTyrLeuGln 88
DB 3116 ACGGCAGATACGTATCTGGAG 3136

RESULT 10
BSP133614 4161 bp DNA linear BCT 09-MAY-2001
LOCUS Bacillus sp. BP23 celB gene.
DEFINITION ACCESSION AJ133614
VERSION AJ133614.1 GI:4490765
KEYWORDS celB gene; cellulase; glycosyl hydrolase.
SOURCE Bacillus sp. BP-23
ORGANISM Bacillus sp. BP-23
REFERENCE 1
AUTHORS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
TITLES Pastor,F.I., Pujol,X., Blanco,A., Vidal,T., Torres,A.L. and Diaz,P.
Molecular cloning and characterization of a multidomain
endoglucanase from Paenibacillus sp BP-23: evaluation of its
performance in pulp refining
JOURNAL Appl. Microbiol. Biotechnol. 55 (1), 61-68 (2001)
MEDLINE 21129642
PUBMED 11234960
REFERENCE 2 (bases 1 to 4161)
AUTHORS Diaz,P.
TITLES Direct Submission
JOURNAL Submitted (10-MAR-1999) Diaz P., Microbiology, University of
Barcelona, Av.Diagonal 645, Barcelona-08028, SPAIN
FEATURES
source
location/Qualifiers
1..4161
/organism="Bacillus sp. BP-23"
/mol_type="genomic DNA"
/strain="BP-23"
/db_xref="taxon:89769"
398..3525
/gene="celB"
-35_signal 398..404
/gene="celB"
-10_signal 437..442
/gene="celB"
RBS 514..519
/gene="celB"
CDS 532..3525

/gene="celB"
/EC_number="3.2.1.4"
/standard_name="endoglucanase"
/codon_start=1
/transl_table=1
/product="cellulase"
/protein_id="CAB38941.1"
/db_xref="GI:4490766"
/db_xref="GOA:Q92411"
/db_xref="SPTREMBL:Q92411"
/translation="MKGSWRRVAIVTLTSLAGLLAGSFMNTGLRQADAAAGDHNYAEA
LOKALYFVEAQRSGPLPANNRVEWRGDSQMOCADYGVDLTGGWYDAGDHVKGFPWA
ASATWLMSVVEYADYEQAGLEBEIKDNIRWATDFPKAHTTPELNLWGQVAGNTDH
ELYNPADTYRGRYPTWTQGWDSKHYGAQILLARIITNLNMPKATKFTQSTERNLDYOT
DWSTSGGSANNPFTWQGWDSKHYGAQILLARIITNLNMPKATKFTQSTERNLDYOT
VCTNGRVKVTYTGGLAWLQWGLSLRYAANAATFISFVYSDWVSPDKKSRVQNATSOI
NYILGNPQSSVYVCGYQNSQHPHRTAHSSWMNEDI PANHRHILYAMVGGPNA
SQOYTDIDGYVSNVATDINAGTGTALAKNLLYGNHQLANFPAFEPVPEVFE
AAVRSGSNTYETIRALLNRRSGWPMQDLSFKYFLDLSEVYAAGRTSDVQVTVSS
SEGATVQPVVDAAKRIYAITANFNTKIPYGGEGNYRKEVOFRITPGQAWNPAND
PSYQNLTTGNPKVNSYIPVYDAGVKVSGQBPVTPVAPAAPAGVQACVAGNSOVALNW
SASAGAVSYTVKRAEVSQGPYTTVAAGVAGLTYTNTGLTNGKTYVYVVTAVNSAGESP
ASVOVSGMPQATTVPGAVTILSGTAGNQNMLSWTAASGASSYTVQRAVAGGTVDVA
TGLAVLYNDDTALNGTSYSYRVAANVNASGTLNVLTVMLTPSGPPVSTGLEVOYRSG
GSGNSNAVTPQNLNLTQATIDLSVIRYFTKDGTEELSFWCDYAOVGSANVG
MFVAVNPAGKTADTYVEISFTSGAGSLAAGATGVIQTRFSKNWNSAFDQSDNYSYDA
SKTAPFAANKNVTAYQNTQVWGLEP"
532..633
sig_peptide
/gene="celB"

ORIGIN
Alignment Scores:
Pred. No.: 4,83e-15 Length: 4161
Score: 204.50 Matches: 38
Percent Similarity: 67.82% Conservative: 21
Best Local Similarity: 43.68% Mismatches: 27
Query Match: 43.51% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x BSP133614 (1-4161)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 3067 ACAGGTACGCTTGAGGTGCAGTATCGAAGCGGGGTTTCAGGCAATTCAGCAATGCAGTT 3126
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 3127 ACTCCGCAATTCATCTGAAGATACCGGCACACAGCGGATTCATCTGAGTACGTGAAA 3186
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 3187 ATTCTGTATTACTTTACCAAGAC---GGCAGCGAGAGCTGCTCTCTGTTGATTAT 3243
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 3244 GCGCAGGTGGCTCAGCGAAGCTACAAAGGAATGTTGTAGCAGTGAATCCGCGAAGGT 3303
QY 82 ThrAlaAspThrTyrLeuGln 88
DB 3304 ACAGCGGATACGTATGTGGAG 3324

RESULT 11
E35142
LOCUS Truncated cellulase composition.
DEFINITION E35142
ACCESSION E35142
VERSION E35142.1 GI:13018967
KEYWORDS JP 1999221086-A/44.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 2029)

```



AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,  
Hyu,M. and Daian,P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 1999221086-A 44 17-AUG-1999;  
CLARIANT INTERNATIONAL LTD  
COMMENT OS Artificial Sequence  
PN JP 1999221086-A/44  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 1998283606  
PR 19-SEP-1997 US 08/932571  
PI PAIJI ANDERSON,PETAER L. BAGUKUISUTO,ROY M DANIEL, PI  
GURAHAMU K PARINTON,  
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONOTISU WILLIAM  
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42// (C12N1/21,C12R1:19), PC  
(C12N9/42,C12R1:19),C12N15/00  
CC Location/Qualifiers  
FH Key 1..2029  
FT source /organism='Artificial Sequence'.  
FT Location/Qualifiers  
FEATURES  
source  
1..2029  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,6e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservative: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1  
US-09-917-376-4 (1-89) x E35142 (1-2029)  
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAGAACATGAGCAAGTGGCGAGCACAGGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGCCGCGTGTGAAGTACTGTACAAGAACATGAGCAAGTGGCGAGTTCTAGCAGGTT 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGATACGTGTACACACAGTGGTGTGACCAAGCCACAGAGTGGCGTA---TGTGAC 177  
QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACAGATAGGGCGAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATTATTACCTGGAG 261  
RESULT 12  
E35143  
LOCUS 2029 bp DNA linear PAT 18-JUN-2001  
DEFINITION Truncated cellulase composition.  
ACCESSION E35143  
VERSION E35143.1 GI:13018968  
KEYWORDS JP 1999221086-A/45.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2029)  
AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,  
Hyu,M. and Daian,P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 1999221086-A 45 17-AUG-1999;  
CLARIANT INTERNATIONAL LTD  
COMMENT OS Unidentified

PN JP 1999221086-A/45  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 1998283606  
PR 19-SEP-1997 US 08/932571  
PI PAIJI ANDERSON,PETAER L. BAGUKUISUTO,ROY M DANIEL, PI  
GURAHAMU K PARINTON,  
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONOTISU WILLIAM  
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42// (C12N1/21,C12R1:19), PC  
(C12N9/42,C12R1:19),C12N15/00  
CC Location/Qualifiers  
FH Key 1..2029  
FT source /organism='Unidentified'.  
FT Location/Qualifiers  
FEATURES  
source  
1..2029  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,6e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservative: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1  
US-09-917-376-4 (1-89) x E35143 (1-2029)  
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAGAACATGAGCAAGTGGCGAGCACAGGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGCCGCGTGTGAAGTACTGTACAAGAACATGAGCAAGTGGCGAGTTCTAGCAGGTT 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGATACGTGTACACACAGTGGTGTGACCAAGCCACAGAGTGGCGTA---TGTGAC 177  
QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACAGATAGGGCGAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATTATTACCTGGAG 261  
RESULT 13  
AF078038S5  
LOCUS 6005 bp DNA linear BCT 11-FEB-2001  
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial cds; and CelE gene, complete cds.  
ACCESSION AF078042  
VERSION AF078042.1 GI:12743878  
KEYWORDS 5 of 5  
SEGMENT Caldicellulosiruptor sp. Tok7B.1  
SOURCE Caldicellulosiruptor sp. Tok7B.1  
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales;  
Syntrophomonadaceae; Caldicellulosiruptor.  
REFERENCE 1 (bases 1 to 6005)  
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,  
Williams,D.P. and Bergquist,P.L.  
TITLE Multidomain and multifunctional glycosyl hydrolases from the  
extreme thermophile Caldicellulosiruptor isolate Tok7B.1  
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)  
MEDLINE 20171169  
PUBMED 10706665  
REFERENCE 2 (bases 1 to 6005)  
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 123.455 Seconds  
(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVQVYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114534\_28231/app\_query.fasta\_1.1429  
-DB=N Geneseq 29Jan04 -QMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_1\_758@runat\_11052004\_114534\_28231 -NCPU=6 -ICPU=3  
-NO WMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001as:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002as:\*
- 7: geneseqn2003as:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	2289	7	ABZ77633 Nucleotid
2	466	99.1	3687	7	ABZ77634 Nucleotid
3	462	98.3	2869	7	ABZ77632 Nucleotid
4	462	98.3	2869	9	ADD22922 Acidothe
5	436	92.8	3365	7	ABZ76162 A. cellu
6	231.5	49.3	2600	2	Aaql5178 Portion o
7	195.5	41.6	2029	2	AAX55660 DNA seque
8	195.5	41.6	2029	6	AAD26568 Active Ce

9	194.5	41.4	6415	2	AAX55662	Aax55662 DNA seque
10	194.5	41.4	6416	6	AAD26526	Aad26526 Active ce
11	192.5	41.0	11706	2	AAX55661	Aax55661 DNA seque
12	192.5	41.0	11707	6	AAD26525	Aad26525 Active ce
13	188	40.0	1434	6	AAL41028	Aal41028 CMCase ge
14	188	40.0	1488	6	AAL41025	Aal41025 CMCase ge
15	188	40.0	2510	6	ABK53202	Abk53202 Bacillus
16	186	39.6	2175	2	AAQ49820	Aaq49820 NK-1 cell
17	182.5	38.8	2977	2	AAQ13001	Aaq13001 Endo1 gen
18	165	35.1	1438	3	AAZ45336	Aaz45336 DNA encod
19	165	35.1	1482	2	AAZ30978	Aaz30978 DNA encod
20	165	35.1	1482	2	AAZ31562	Aaz31562 Pectate l
21	165	35.1	5562	2	AAT86625	Aat86625 C. thermo
22	157	33.4	1314	6	ABK73393	Abk73393 Bacillus
23	155.5	33.1	4059	5	AAH75059	Aah75059 Nucleotid
24	150.5	32.0	4056	5	AAH75060	Aah75060 Nucleotid
25	130.5	27.8	4452	7	ACA27429	Aca27429 Prokaryot
26	126.5	26.9	1303	9	ADC27473	Adc27473 Fusion pr
27	126.5	26.9	1747	9	ADC27474	Adc27474 Fusion pr
28	126.5	26.9	3489	9	ADC27475	Adc27475 Fusion pr
29	120.5	25.6	1775	2	AAQ13003	Aaq13003 Endo3 gen
30	114	24.3	486	2	AAQ72917	Aaq72917 Cellulose
31	114	24.3	486	2	AAV74072	Aav74072 C. cellul
32	114	24.3	499	2	AAZ4930	Aaz4930 Clostridi
33	114	24.3	507	5	AAD11042	Aad11042 Clostridi
34	114	24.3	768	6	AAF86248	Aaf86248 DNA seque
35	114	24.3	768	6	ABK52403	Abk52403 E. coli c
36	114	24.3	984	5	AAD11046	Aad11046 Chimeric
37	114	24.3	1030	5	AAD11044	Aad11044 Clostridi
38	114	24.3	1288	5	AAD11045	Aad11045 Clostridi
39	113	24.0	573	5	AAD11043	Aad11043 Clostridi
40	113	24.0	1146	2	AAZ27550	Aaz27550 Gaussia l
41	113	24.0	1146	6	AAD22201	Aad22201 Gaussia s
42	78.5	16.7	764	3	AAI12881	Aai12881 Aspergill
43	78.5	16.7	3147	7	ACA25846	Aca25846 Prokaryot
44	77	16.4	8107	3	AAZ92616	Aaz92616 Genomic D
45	76	16.2	96597	8	ADA02720	Ada02720 Human NFA

ALIGNMENTS

RESULT 1	ABZ77633	standard; DNA; 2289 BP.
ID	ABZ77633	standard; DNA; 2289 BP.
XX	XX	
AC	ABZ77633;	
XX	XX	
DT	03-JUN-2003 (first entry)	
XX	XX	
DE	Nucleotide sequence of the ManA polypeptide.	
XX	XX	
KW	ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;	
KW	food; feed; paper pulp; biofuel; mannanase; gene; ss.	
XX	XX	
OS	Acidothermus cellulolyticus.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	1..2289
FT	FT	/*tag= a
FT	FT	/product= "ManA"
XX	XX	
PN	WO2003012110-A1.	
XX	XX	
PD	13-FEB-2003.	
XX	XX	
PF	28-JUL-2001; 2001WO-US023819.	
XX	XX	
PR	28-JUL-2001; 2001WO-US023819.	
XX	XX	
PA	(MIDE ) MIDWEST RES INST.	
XX	XX	
PI	Ding S, Adney WS, Vinzant TB, Himmel ME;	

```

DR WPI; 2003-248182/24.
DR P-PSDB; ABP73022.
XX
XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
XX Example 1; Page 23; 46pp; English.
XX
XX The present sequence encodes ManaA, a thermostable mannanase A polypeptide
CC derived from Acidothermus cellulolyticus. ManaA is a member of the
CC glycoside hydrolase family of enzymes. ManaA is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. ManaA is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying ManaA, or detecting ManaA polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of ManaA
CC polypeptides
XX
XX Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;
XX
XX Alignment Scores:
Pred. No.: 4.13e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
XX
XX US-09-917-376-4 (1-89) x ABZ77633 (1-2289)
XX
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCCGGGTGATACACG 1422
XX
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGTGGATTGTTCGACGGTG 1482
XX
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 1483 ACGGTGCGGTACTGTGTTTACCCGGGATGGTGGTCTGTCGACACTGCTGTACAACTGTGAC 1542
XX
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 1543 TGGGGCGCGATGGGGTGGGAATATCCGCGCTCTGTTCCGCTCGGTGAACCCGGCGACG 1602
XX
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 1603 CCGACGGCGGACACCTACCTGCAG 1626
XX
RESULT 2
ABZ77634
ID ABZ77634 standard; DNA; 3687 BP.
XX
XX AC ABZ77634;
XX
XX 03-JUN-2003 (first entry)
XX
XX Nucleotide sequence of the GuxA polypeptide.
XX
XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW gene; ss.
XX
XX Acidothermus cellulolyticus.
XX
XX Key Location/Qualifiers
FT CDS 1..3687
FT /tag= a
FT /product= "GuxA"

```

---

```

XX WO2003012109-A1.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001WO-US023817.
XX
XX 28-JUL-2001; 2001WO-US023817.
XX
XX (MIDE ) MIDWEST RES INST.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
XX WPI; 2003-239526/23.
XX
XX P-PSDB; ABP73029.
XX
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing cellulose in a starting material, and
PT for the conversion of biomass to biofuels and biofuel additives.
XX
XX Example 1; Page 23-24; 47pp; English.
XX
XX The present sequence encodes a GuxA polypeptide. GuxA is thermostable
CC cellulase, and is a member of the glycoside hydrolase family of enzymes.
CC GuxA is useful for reducing cellulose in a starting material such as
CC agricultural biomass to sugars. This is useful in biofuel production.
CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
CC additives, in detergents, pulp and paper processing, food and feed
CC processing, and in textile process. GuxA is also useful for raising
CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
CC or detecting GuxA polypeptide expression, as well as reagent tools for
CC characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
XX
XX Alignment Scores:
Pred. No.: 1.79e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
DB: 7 Gaps: 0
XX
XX US-09-917-376-4 (1-89) x ABZ77634 (1-3687)
XX
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCCGGGTGATAACACG 1809
XX
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGTGGATTGTTCGACGGTG 1869
XX
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 1870 ACGGTGCGGTACTGTGTTTACCCGGGATGGTGGTCTGTCGACACTGCTGTCAACTGTGAC 1929
XX
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 1930 TGGGGCGCGATGGGGTGGGAATATCCGCGCTCTGTTCCGCTCGGTGAACCCGGCGACG 1989
XX
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 1990 CCGACGGCGGACACCTACCTGCAG 2013
XX
XX RESULT 3
XX ABZ77632
XX ID ABZ77632 standard; DNA; 2869 BP.
XX
XX AC ABZ77632;
XX
XX 03-JUN-2003 (first entry)
XX
XX

```



```

DB:          9          Gaps:          0
US-09-917-376-4 (1-89) x ADD22922 (1-2869)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGTGGGTGAAGTGCAGTATAAGATATGATTCGGCGCGGTGATATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTTCGAGGTGGTGAATACCGGGTCGTGCGTGGATTGTGCGACGGT 2724
QY 41 ThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGTGTTCACCGGGATGGTGGCTGCTGCGACACTGGTGACACTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGGGGATCGGGTGGGATATCCGGCTCGTTCGGTTCGGTGAACCGGGCGGACG 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868
RESULT 5
ABZ76162
ID ABZ76162 standard; DNA; 3365 BP.
XX AC ABZ76162;
XX DT 29-MAY-2003 (first entry)
XX DE A. cellulolyticus Gux1 protein encoding DNA.
KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;
KW biofuel; detergent; pulp; paper processing; feed processing; textile;
KW cellulase; gene; ds.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
XX FT CDS 1..3365
XX FT /product= "Gux1"
XX FT /transl_except= (pos: 682..683, aa: Pro)
XX FT /note= "this codon has an apparent one nucleotide
XX FT basepair deletion which alters the reading frame"
XX PN WO2003012095-A1.
XX XX 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023820.
XX PR 28-JUL-2001; 2001WO-US023820.
XX (MIDE ) MIDWEST RES INST.
XX PA Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;
XX PI Lantz Mccarter S;
XX XX WPI; 2003-300494/29.
XX DR P-PSDB; ABP71656.
XX XX New thermal tolerant Gux1 peptide having specified amino acid sequence,
XX PT useful in the degradation of cellulose to biofuels.
XX PS Disclosure; Page 22-23; 44pp; English.
XX CC The invention relates to a thermal tolerant Gux1 peptide from A.
XX CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside
XX CC hydrolase family and comprises a catalytic domain GH48, carbohydrate
XX CC binding domain type III, and a carbohydrate binding domain type II. The

```

```

CC polypeptide is useful in the degradation of cellulose into biofuel, or
CC for conversion of biomass to biofuel additives. It is used in detergents,
CC pulp and paper processing, food and feed processing, and in textile
CC processing. It can also be used alone or in combination with other
CC cellulase or glycoside hydrolases. The novel polypeptide generates
CC alternative cellulase enzymes capable of assisting in the commercial-
CC scale processing of cellulose to sugar for use in biofuel production. The
CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding
XX DNA
SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      8.56e-44      Length:      3365
Score:          436.00      Matches:      81
Percent Similarity: 97.62%      Conservative: 1
Best Local Similarity: 96.43%      Mismatches: 2
Query Match:      92.77%      Indels:      0
DB:              7          Gaps:      0

US-09-917-376-4 (1-89) x ABZ76162 (1-3365)
QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 112 CTCAAAGCGCAGTATAGAACAAATGATTCGGCGCGAGTGACACACAGATCAACCGGGT 171
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 172 CTCACGTGTGGTGAATACCGGGTCGTGCTGGTGGATTGTCGACGGTGACGGTGCATG 231
QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspThrAlaAlaMet 64
Db 232 TGGTTTCAACCGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 291
QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84
Db 292 GGGTGTGGGAATATCCGGCTCGTTCGGTGAACCCGGCGACCCCGCGCGGAC 351
QY 85 ThrTyrLeuGln 88
Db 352 ACCTACTGCAG 363
RESULT 6
AAQ15178
ID AAQ15178 standard; DNA; 2600 BP.
XX AC AAQ15178;
XX XX 27-AUG-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 17-MAR-1992 (first entry)
XX DE Portion of pAEC-1 contg. cellulase AE-1 gene.
XX KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.
XX OS Aeromonas sp.
XX FH Key Location/Qualifiers
XX FT CDS 192..2540
XX FT /*tag= a
XX FT /product= "cellulase AE-1"
XX PN JP03251174-A.
XX XX 08-NOV-1991.
XX PF 28-FEB-1990; 90JP-00045465.
XX PR 28-FEB-1990; 90JP-00045465.
XX PA (OJIP ) OJI PAPER CO.
XX XX

```



DR WPI; 1991-373412/51.  
DR P-PSDB; AAR15625.  
XX  
XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
PT pH when carboxymethylcellulose is used as substrate.  
XX  
XX Claim 3; Fig 2; 8pp; Japanese.  
PS  
XX The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene  
CC (obcd. from Aeromonas) into pUC18. The gene can be used to produce  
CC recombinant enzyme which is used for the effective utilis- ation of  
CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and  
CC also for the detergent and deinking of waste paper. (Updated on 25-MAR-  
XX 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:  
Pred. NO.: 1.62e-18 Length: 2600  
Score: 231.50 Matches: 42  
Percent Similarity: 66.67% Conservative: 16  
Best Local Similarity: 48.28% Mismatches: 28  
Query Match: 49.28% Indels: 1  
DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAQ15178 (1-2600)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 2082 AGCGGGGATTTGGCGTTTCAGTACAAAACCGGAGATGCAATGCGCCGACACAGTTC 2141

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 2142 AAGCTCTATTCAATATCGTGAACGAGGGCTGGCGAGTGCCTTTAAGCGAGCTGAGT 2201

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 2202 CTGCGATATTATTTTCACAGCTGAC---GGCAATGACCAACTGCAATACAAATTTGCTACTGG 2258

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 2259 GCCATGTGTAGGTCTCAACCTGAACGGGGCTTTTCGTGAAAATGAATCCGGGCAAGCG 2318

QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 2319 AACGCCGATACCTACTTGGAG 2339

RESULT 7  
ID AAX55660  
XX AAX55660 standard; DNA; 2029 BP.  
AC AAX55660;  
XX  
DT 30-JUL-1999 (first entry)  
XX  
DE DNA sequence encoding truncated cellulase Cel E3/B5.  
XX  
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
KW cotton-containing fabric; stonewashing; ss.  
XX  
OS Unidentified.  
XX  
PN EP921188-A2.  
XX  
PD 09-JUN-1999.  
XX  
PF 15-SEP-1998; 98EP-00810919.  
XX  
PR 19-SEP-1997; 97US-00932571.  
XX  
PA (CLRN ) CLARIANT FINANCE BVI LTD.  
XX

PI Anderson P., Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
PI Morgan H., Williams DP;  
XX  
XX WPI; 1999-315403/27.  
DR P-PSDB; AAV13494.  
XX  
XX New truncated cellulase proteins, useful in detergents and for producing  
PT 'stonewashed' denim.  
XX  
XX Disclosure; Page 41-42; 65pp; English.  
XX  
XX The invention relates to a recombinant cellulase active protein free of  
CC proteinases of native thermophilic and alkalophilic origin, comprising  
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,  
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
CC length sequences, or functional equivalents. Cel B5 extends from amino  
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
CC and the stability region extends from amino acid E482 to G635 in the  
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
CC enzymes are useful in laundry detergent compositions to prevent or remove  
CC staining, backstaining or graying, for use on cellulosic materials  
CC including cotton-containing fabrics. They are especially useful for  
CC preventing redeposition of colorant during stonewashing, and for  
CC processing of textiles where cellulose breakdown is required. The new  
CC truncated enzymes show reduced redeposition of dye compared to using non-  
CC truncated cellulase compositions  
XX

SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:  
Pred. NO.: 3.53e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservative: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAX55660 (1-2029)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGGTGTGAAGGTACTGTACAGACAAATGAGACAGTGCAGCACAGTTCT 60

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGGCCGCTGGTTTAAGATAGTAGTAATGGAGGCGAGCAGTGTGTGATCTTAGCAGGTT 120

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGATACCTGTGTACACAGTGGATGGTGTGACAGCCACAGAGTGCAGTA---TGTGAC 177

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACATAGGGCCCAAGCAATGACATTCATTTTGTGAGCTTAGCAGCGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATATTACCTGGAG 261

RESULT 8  
ID AAD26568  
XX AAD26568 standard; DNA; 2029 BP.  
AC AAD26568;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Active cellulase hybrid protein, E3/B5 DNA.  
XX  
KW Active cellulase protein; alkalophilic; textile processing; proteinase;



QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
 DB 2664 GGTGTGAAGTACTGTCAAGAACAAATGAGACAAGTGGCGACACAGGTTCTATAAGGCG 2723  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 DB 2724 TGGTTTAAGATAGTAAGTGGAGGACGACAGTGTGATCTTAGCAGGTTAAGATAAGA 2783  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 DB 2784 TACTGGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2840  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 DB 2841 ATAGGGCAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2900  
 QY 84 AspThrTyrLeuGln 88  
 DB 2901 GATTATTACCTGGAG 2915  
 RESULT 10  
 AAD26526  
 ID AAD26526 standard; DNA; 6416 BP.  
 AC AAD26526;  
 DT 26-MAR-2002 (first entry)  
 DE Active cellulase protein, cele gene.  
 XX  
 KW Active cellulase protein; alkalophilic; textile processing; proteinase;  
 KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; cele gene; ds.  
 XX  
 OS Unidentified.  
 FH  
 FT Key Location/Qualifiers  
 FT CDS 634..5889  
 FT /\*tag= a  
 FT /product= "cele protein"  
 FT misc\_feature 748..2538  
 FT /\*tag= c  
 FT /product= "DNA encoding E1/2 protein"  
 FT misc\_feature 748..2076  
 FT /\*tag= b  
 FT /product= "DNA encoding E1 protein"  
 XX  
 PN US6294366-B1.  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 19-AUG-1998; 98US-00136574.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 DR WPI; 2002-081780/11.  
 DR P-PSDB; AAEL6324.  
 XX  
 PT New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 PS Disclosure; Col 37-44; 6lpp; English.  
 XX  
 CC The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the

CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is cele gene  
 XX  
 SQ Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;  
 Alignment Scores:  
 Pred. No.: 2,05e-13 Length: 6416  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservative: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.38% Indels: 1  
 DB: 6 Gaps: 1  
 US-09-917-376-4 (1-89) x AAD26526 (1-6416)  
 QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
 DB 2665 GGTGTGAAGTACTGTCAAGAACAAATGAGACAAGTGGCGACACAGGTTCTATAAGGCG 2724  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 DB 2725 TGGTTTAAGATAGTAAGTGGAGGACGACAGTGTGATCTTAGCAGGTTAAGATAAGA 2784  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 DB 2785 TACTGGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2841  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 DB 2842 ATAGGGCAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2901  
 QY 84 AspThrTyrLeuGln 88  
 DB 2902 GATTATTACCTGGAG 2916  
 RESULT 11  
 AAX55661  
 ID AAX55661 standard; DNA; 11706 BP.  
 AC AAX55661;  
 XX  
 DT 30-JUL-1999 (first entry)  
 XX  
 DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.  
 XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN EP921188-A2.  
 XX  
 PD 09-JUN-1999.  
 XX  
 PF 15-SEP-1998; 98EP-00810919.  
 XX  
 PR 19-SEP-1997; 97US-00932571.  
 XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 DR WPI; 1999-315403/27.  
 DR P-PSDB; AAV13492.

```

XX New truncated cellulase proteins, useful in detergents and for producing
PT 'stonewashed' denim.
XX
XX Disclosure; Page 20-23; 65pp; English.
XX
XX The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC extends from Y39 to G812, Cel B6 extends from amino acid V1233 to K1751
CC and the stability region extends from amino acid E482 to G635 in the
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or remove
CC staining, backstaining or graying, for use on cellulosic materials
CC including cotton-containing fabrics. They are especially useful for
CC preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using non-
CC truncated cellulase compositions
XX
SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,82e-13 Length: 11706
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservatives: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 40.96% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAX55661 (1-11706)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 4038 GGTTCGAAGTACTATACAAAGAACATGAGCAAGTCGCGAGCAAGTTCTATAAGCGCG 4097
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 4098 TGGTTTAAAGATAGTGAATGAGGAGCAGCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGA 4157
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63
DB 4158 TACTGTGACACAGTGGATGGTGACAGGCCAGAGAGTGGCGTA---TGTGACTGGGCGACAG 4214
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 4215 ATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTGAGCGAGCGAGTGGAGCG 4274
QY 84 AspThrTyrLeuGln 88
DB 4275 GATTATTACTTGGAG 4289

RESULT 12
AAD26525
ID AAD26525 standard; DNA; 11707 BP.
XX
XX AAD26525;
XX
XX 26-MAR-2002 (first entry)
XX
XX Active cellulase protein, celB gene.
XX
XX Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; celB gene; ds.
XX
XX Unidentified.
OS
XX

```

```

FH Key Location/Qualifiers
CDS 6255..10535
FT /*tag= a
FT /product= "CelB protein"
FT misc_feature 8501..10532
FT /*tag= b
FT /product= "DNA encoding B4/5 protein"
FT misc_feature 9255..10526
FT /*tag= c
FT /product= "DNA encoding B/5 protein"
XX
XX US6294366-B1.
XX
XX 25-SEP-2001.
XX
XX 19-AUG-1998; 98US-00136574.
XX
XX 19-SEP-1997; 97US-00932571.
XX
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
XX Morgan H, Williams DP;
XX
XX WPI; 2002-081780/11.
XX P-PSDB; AAE16323.
XX
XX New cellulase active protein, useful in textile processing or commercial
XX detergents, e.g. for improving the feel or appearance of cotton-
XX containing fabrics, is stable under conditions of alkaline pH and
XX elevated temperatures.
XX
XX Disclosure; Col 27-38; 61pp; English.
XX
XX The present invention relates to a cellulase active protein, which is
XX substantially free of proteinases of native thermophilic and
XX alkaliphilic origin, where the cellulase active protein consists of the
XX CelB5 amino acid sequence. The cellulase active protein is useful for
XX treating cellulosic materials including cotton-containing fabrics, as
XX detergent additives. The cellulase active protein is also useful for
XX improving the feel and/or appearance of cotton-containing fabrics, for
XX removing surface fibers from cotton-containing knits or for imparting
XX stonewashed appearance to cotton-containing denims. The present proteins
XX are stable under condition of alkaline pH and elevated temperatures, thus
XX suitable for textile processing and in commercial detergents. The present
XX sequence is celB gene
XX
SQ Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,83e-13 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservatives: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 40.96% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x AAD26525 (1-11707)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 4038 GGTTCGAAGTACTATACAAAGAACATGAGCAAGTCGCGAGCAAGTTCTATAAGCGCG 4097
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 4098 TGGTTTAAAGATAGTGAATGAGGAGCAGCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGA 4157
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63
DB 4158 TACTGTGACACAGTGGATGGTGACAGGCCAGAGAGTGGCGTA---TGTGACTGGGCGACAG 4214
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

```

Db	4215	ATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTGACGACGGAGTGAGTGGACGG	4274
QY	84	AspThrTyrLeuGln 88	
Db	4275	GATTATTACTTGGAG	4289
RESULT 13			
ID	AAL41028		
XX	AAL41028 standard; DNA; 1434 BP.		
AC	AAL41028;		
XX	11-OCT-2002 (first entry)		
DE	CMCase gene with additional His encoding sequences.		
KW	Protein surface-displayed; genetic carrier; spore; monoclonal variant;		
KW	high-throughput screening; antibody; vaccine; enzyme; gene; ds.		
OS	Unidentified.		
XX	Key	Location/Qualifiers	
FT	CDS	1..1434	
FT	/*tag= a		
FT	/product= "CMCase protein"		
XX	WO200255561-A1.		
XX	18-JUL-2002.		
XX	15-JAN-2002; 2002WO-KR000059.		
XX	15-JAN-2001; 2001KR-00002156.		
PA	(GENO-) GENOFOCUS CO LTD.		
PI	Pan JG, Choi SK, Jung HC;		
DR	WPI; 2002-590661/63.		
DR	P-PSDB; AAO22444.		
XX	Methods for surface display of proteins, e.g. displaying a protein on the surface of spores, improving a protein or isolating a substance, by transforming a host cell harboring a genetic carrier with spore or virus with the vector library.		
XX	Example 5; Page 109-113; 118pp; English.		
CC	The invention relates to methods for preparing a protein surface-		
CC	displayed on a genetic carrier, improving a protein or isolating a		
CC	substance in a mixture comprising transforming a host cell harbouring a		
CC	genetic carrier consisting of a spore or virus with the vector library.		
CC	The method is useful for surface display of proteins, particularly for		
CC	displaying a protein on the surface of e.g. spores, for improving a		
CC	protein, or for isolating a substance. These methods are useful in		
CC	obtaining monoclonal variants from a large library, or in high-throughput		
CC	screening of antibodies for use in therapy (e.g. as vaccines), diagnosis		
CC	or analysis. This polynucleotide sequence represents the CMCase gene with		
CC	additional His encoding sequences relating to the invention		
XX	Sequence 1434 BP; 493 A; 293 C; 341 G; 307 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1.94e-13	Length:	1434
Score:	188.00	Matches:	34
Percent Similarity:	61.18%	Conservative:	18
Best Local Similarity:	40.00%	Mismatches:	31
Query Match:	40.00%	Indels:	2
DB:	6	Gaps:	1
US-09-917-376-4 (1-89) x AAL41028 (1-1434)			
QY	4	GlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23

Db	991	GGTATTTCGTCAATACAGACGAGGGGATGGAGTATGAACAGCAACCAATCCGTCCG	1050
QY	24	GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg	43
Db	1051	CAGCTTCAATAAATAACGCAATACCAACGGTTGATTTTAAAGATGTCACTGCCGT	1110
QY	44	TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	1111	TACTGTATACGCGAAAAACAAAGCCAAACGTT-----GACTGTGACTACGCGAG	1164
QY	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
Db	1165	CTTGGATCGGCAATGTGACATACAAGTTTGTGACGTGTGCATTAACCAACCAAGGTGCA	1224
QY	84	AspThrTyrLeuGln 88	
Db	1225	GATACCTATCTGGAA	1239
RESULT 14			
ID	AAL41025		
XX	AAL41025 standard; DNA; 1488 BP.		
AC	AAL41025;		
XX	11-OCT-2002 (first entry)		
DE	CMCase gene with mutated signal sequence to enhance hydrophobicity.		
XX	Protein surface-displayed; genetic carrier; spore; monoclonal variant;		
KW	high-throughput screening; antibody; vaccine; enzyme; gene; ds.		
OS	Unidentified.		
XX	Key	Location/Qualifiers	
FT	CDS	1..1491	
FT	/*tag= a		
FT	/product= "CMCase protein"		
XX	WO200255561-A1.		
XX	18-JUL-2002.		
XX	15-JAN-2002; 2002WO-KR000059.		
XX	15-JAN-2001; 2001KR-00002156.		
PA	(GENO-) GENOFOCUS CO LTD.		
PI	Pan JG, Choi SK, Jung HC;		
DR	WPI; 2002-590661/63.		
DR	P-PSDB; AAO22443.		
XX	Methods for surface display of proteins, e.g. displaying a protein on the surface of spores, improving a protein or isolating a substance, by transforming a host cell harboring a genetic carrier with spore or virus with the vector library.		
XX	Example 4; Page 98-103; 118pp; English.		
CC	The invention relates to methods for preparing a protein surface-		
CC	displayed on a genetic carrier, improving a protein or isolating a		
CC	substance in a mixture comprising transforming a host cell harbouring a		
CC	genetic carrier consisting of a spore or virus with the vector library.		
CC	The method is useful for surface display of proteins, particularly for		
CC	displaying a protein on the surface of e.g. spores, for improving a		
CC	protein, or for isolating a substance. These methods are useful in		
CC	obtaining monoclonal variants from a large library, or in high-throughput		
CC	screening of antibodies for use in therapy (e.g. as vaccines), diagnosis		
CC	or analysis. This polynucleotide sequence represents the CMCase gene with		
CC	a mutated signal sequence to enhance its hydrophobicity relating to the		
CC	invention		

```
XX SQ Sequence 1488 BP; 501 A; 297 C; 357 G; 333 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,04e-13 Length: 1488
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x AAL41025 (1-1488)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1048 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGACAGCAACCAATCCGTCGG 1107
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1108 CAGCTTCAATAAATAAATAACGCAATACCGGTTGATTAAAGATGCTACTGCCGT 1167
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 1168 TACTGGTATACCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGGCAG 1221
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1222 CTTGGATCGGCAATGTGCATACAAAGTTTGTGACGTTCATAAACCAAGCAAGGTGCA 1281
QY 84 AspThrTyrLeuGln 88
Db 1282 GATACCTATCTGGAA 1296

RESULT 15
ID ABK53202 standard; DNA; 2510 BP.
XX AC ABK53202;
XX DT 12-AUG-2002 (first entry)
DE Bacillus subtilis spore coat protein cotG-CMCCase fusion gene.
XX KW Spore coat protein; spore surface; delivery system; bioconversion;
KW recombinant live vaccine; resistance property; protein microarray;
KW whole cell absorber; secretion signal; target signal; cell membrane;
KW surface display motif; cotG; CMCCase; carboxymethyl cellulase; gene; ds.
XX OS Bacillus subtilis.
OS Synthetic.
OS Chimeric.
XX PH Key
FT promoter 1..460 Location/Qualifiers
FT /tag= a
FT /note= "CotG promoter"
FT CDS 461..2494
FT /tag= b
FT /partial
FT /product= "Bacillus subtilis spore coat protein cotG-
FT carboxymethyl cellulase (CMCCase) fusion protein"
FT /note= "This sequence lacks a start codon"
FT misc_feature 461..1045
FT /tag= c
FT /note= "CotG structural gene"
FT misc_feature 1046..1084
FT /tag= d
FT /note= "Linker"
FT misc_feature 1085..2491
FT /tag= e
FT /note= "CMCCase structural gene"
XX PN W0200246388-A1.
```

```
XX 13-JUN-2002.
PD 07-DEC-2001; 2001WO-KR002124.
XX PF
XX PR 08-DEC-2000; 2000KR-00074835.
XX PA (GENO-) GENOFOCUS CO LTD.
XX PI Pan J, Choi S, Jung H;
XX WPI; 2002-463830/49.
DR P-PSDB; AAU98063.
XX
PT Displaying a protein of interest on spore surface, by transforming a host
PT cell with a vector comprising a gene encoding a spore coat protein and a
PT gene encoding a protein of interest, useful for preparing protein
PT microarray.
XX
PS Example 4; Page 145; 158pp; English.
XX
CC The present invention relates to a new method for displaying a protein of
CC interest on a spore surface. The method of the invention involves
CC preparing a vector comprising a gene construct containing a gene encoding
CC a spore coat protein and a gene encoding a protein of interest,
CC transforming a host cell with the vector, displaying the protein of
CC interest on a surface of a spore of the host cell, and recovering the
CC spore. The method is useful for producing a delivery system for
CC recombinant live vaccine. The method is also useful for improving a
CC protein of interest using a resistance property of spore, in
CC bioconversion reactions, for preparing protein microarray, for producing
CC an antibody to antigen in vertebrates, and for preparing a whole cell
CC absorber. The spore coat proteins used circumvent a necessity for passage
CC across cell membrane, so that they do not need secretion signal and
CC target signal which are prerequisites of surface display motif, therefore
CC ensuring a surface display of protein, such as beta-galactosidase, in
CC orderly fashion, which is difficult to pass across cell membrane. The
CC present nucleic acid sequence encodes the Bacillus subtilis spore coat
CC protein cotG-carboxymethyl cellulase (CMCCase) fusion protein of the
CC invention
XX SQ Sequence 2510 BP; 869 A; 528 C; 521 G; 592 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3.97e-13 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x ABK53202 (1-2510)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGACAGCAACCAATCCGTCGG 2110
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCAATAAATAAATAACGCAATACCGGTTGATTAAAGATGTCCTGCCGT 2170
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 2171 TACTGGTATACCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGGCAG 2224
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTTGGATCGGCAATGTGCATACAAAGTTTGTGACGTTCATAAACCAAGCAAGGTGCA 2284
QY 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299
```

Search completed: May 12, 2004, 05:39:24  
Job time : 134.455 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 835.94 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQYKNDSPAGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_5334@runat\_11052004\_114535\_28249 -NCPUL=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_est3.\*  
12: gb\_est4.\*  
13: gb\_est5.\*  
14: gb\_est6.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_nam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rtd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	81	17.2	584	14	CF480372	POL1_65.H
2	79.5	16.9	754	12	BQ047524	BQ047524 EST586642
3	78.5	16.7	768	14	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	13	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	9	AA043716	AA043716 zk61f10.s
6	76	16.2	606	10	BE362891	BE362891 DGI_90.C1
7	76	16.2	641	14	CD226529	CD226529 CCI_46.E
8	76	16.2	657	10	BE362820	BE362820 DGI_89.C0
9	76	16.2	676	14	CF487649	CF487649 POL1_45.C
10	76	16.2	691	12	BG739392	EM1_82.D0
11	76	16.2	724	28	BZ616323	BZ616323 iG58e08.g
12	76	16.2	814	12	BI218457	BI218457 602937951
13	76	16.2	905	28	BZ545807	BZ545807 OGALR72TM
14	76	16.2	934	29	CG270741	CG270741 OGMIB33TH
15	76	16.2	1001	29	CG270748	CG270748 OGMIB33TV
16	75.5	16.1	657	10	BF513128	BF513128 UI-H-BW1
17	75.5	16.1	1058	14	CK231562	CK231562 ILLUMIGEN
18	74.5	15.9	465	10	AW491464	AW491464 UI-M-BH3
19	74.5	15.9	613	14	CA595709	CA595709 wpaic.pk0
20	74.5	15.9	650	28	BZ894843	BZ894843 Hg4_0131
21	74	15.7	302	14	CF076729	CF076729 QHK12L22.
22	74	15.7	653	12	BI664037	BI664037 603289338
23	73.5	15.6	261	10	BF875908	BF875908 RC3-ET013
24	73.5	15.6	816	14	CB293980	CB293980 UCRCS01_0
25	73.5	15.6	819	14	CB293084	CB293084 UCRCS01_0
26	73.5	15.6	1039	12	BG334904	BG334904 602461219
27	73.5	15.6	1237	13	BQ065850	BQ065850 AGENCOURT
28	73	15.5	357	14	CF078232	CF078232 QHK1J23.Y
29	73	15.5	360	13	C64738	C64738 C64738 Yuj1
30	73	15.5	593	14	CF622108	CF622108 lafl3d06.
31	73	15.5	776	29	BX209426	BX209426 Danilo rer
32	73	15.5	861	28	BH319803	BH319803 CH230-115
33	73	15.5	894	29	CG077853	CG077853 PUFKJ89TB
34	73	15.5	924	29	CG077856	CG077856 PUFKJ89TD
35	72.5	15.4	380	12	BG816375	BG816375 dad35808.
36	72.5	15.4	450	9	AL799805	AL799805 AL799805
37	72.5	15.4	533	10	AW963120	AW963120 EST375193
38	72.5	15.4	544	9	AL898643	AL898643 AL898643
39	72.5	15.4	563	28	BZ424716	BZ424716 100018538
40	72.5	15.4	567	14	CF754312	CF754312 EST-77-2-
41	72.5	15.4	579	9	AL962576	AL962576 AL962576
42	72.5	15.4	601	29	CE253602	CE253602 tigr-gss-
43	72.5	15.4	640	9	AL846493	AL846493 AL846493
44	72.5	15.4	648	12	BM609423	BM609423 170006871
45	72.5	15.4	719	12	BG635313	BG635313 AT31857.5

ALIGNMENTS

RESULT 1  
CF480372  
LOCUS  
DEFINITION  
CF480372 POL1\_65\_H07.b1\_A002 Pollen Sorghum bicolor cDNA clone  
ACCESSION  
CF480372 POL1\_65\_H07\_A002 3', mRNA sequence.  
VERSION  
CF480372.1 GI:34509241  
KEYWORDS  
EST.  
SOURCE  
Sorghum bicolor (sorghum)  
ORGANISM  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

```

REFERENCE 1 (bases 1 to 584)
AUTHORS   Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
          Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,
          Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and
          Pratt,L.H.
          EST database from Sorghum: pollen
          Unpublished (2003)
          Other ESTs: POL1_65.H07.g1.A002
          Contact: Cordonnier-Pratt MM
          Laboratory for Genomics and Bioinformatics
          The University of Georgia, Department of Plant Biology
          Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
          Tel: 706 542 1860
          Fax: 706 593 0210
          Email: mmpratt@uga.edu
          Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
          the Human Genome Center, University of Tokyo Institute of Medical
          Science; plant material and RNA prepared at Texas A & M University;
          sequencing done in the laboratory for Genomics and Bioinformatics,
          University of Georgia. Sequence ends have been trimmed to exclude
          vector and regions below Phred quality 16. Three-prime sequences
          are presented as their reverse complement and have been trimmed to
          exclude polyA.
          Seq primer: Sug3-14 (TAGCTAGCGCGCGGACC)
          POLYA=yes.

FEATURES             source
    Location/Qualifiers
        1..584
            /organism="Sorghum bicolor"
            /mol_type="mRNA"
            /cultivar="BTx623"
            /db_xref="taxon:4558"
            /clone="POL1_65.H07.A002"
            /lab_host="DH10B-T1 phage-resistant E. coli"
            /clone_lib="pollen"
            /note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
            Site 2: XhoI; The library was prepared from polyA+ RNA
            from pollen at the late vacuolated-vacuolated stage of
            development. Pollen was harvested from greenhouse-grown
            panicles of sorghum line BTx623. Panicles were removed
            from the flag leaf prior to emergence, when no detectable
            amylase is present in pollen of male-fertile lines. This
            stage represents pollen collected from anthers about 8-14
            days prior to anthesis. Double-stranded cDNA was cloned
            unidirectionally into different DraIII sites of the
            pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG,
            3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA
            insert."

ORIGIN
Alignment Scores:
Pred. No.:      12.7      Length:      584
Score:          81.00      Matches:      24
Percent Similarity: 45.35%      Conservative: 15
Best Local Similarity: 27.91%      Mismatches: 37
Query Match:      17.23%      Indels:    10
DB:              14         Gaps:       4

US-09-917-376-4 (1-89) x CF480372 (1-584)

QY      5  ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
      :::
      45  ATAGCAATCAGCGCAAACTCAAATGAATCCGTCGGGAGGAGCCTGTGCTGCGCGGCG 104
      ::::

QY      25  LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg--- 43
      ::::

DB      105  -----CGCGCGCGGACAGAACACCTTCAAAACGATCGTGCAGAGAAA 152
      ::::

QY      44  ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
      ::::

DB      153  CCATGGTGGGGATCTGCGGGAAGAGGTTTCAGCACCATGGAATATATATATATGATGAA 212
      ::::

QY      62  AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
      ::::

```

---

```

Db      213  GCTTGTGTGAAGTGTGGTGTGAGCTTGTGTTCTTTGAGACGGTGTCTTGTGTTGGTCCG 272

QY      79  AlaThrProThrAlaAsp 84
      ::::

Db      273  AGCTTCAGTCTCGTGAT 290

RESULT 2
BQ047524
LOCUS     BQ047524
DEFINITION Solanum tuberosum (potato)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          asterids; lamiids; Solanales; Solanaceae; Solanum.
          1 (bases 1 to 754)
AUTHORS   Zhang,P., Hernandez,M., Tornqvist,C.-E., Wirtz,U., Loukoianov,A.,
          Rangel,P., Haberlach,G.T., Karamycheva,S.A., Tsai,J., Chieningo,A.,
          Bougri,O., Buell,C.R., Ronning,C.M., Heigeson,J. and Baker,B.
          Generation of ESTs from Potato Leaves Challenged with Phytophthora
          infestans, incompatible Interaction (2002)
          Unpublished (2002)
          Contact: Robin Buell
          The Institute for Genomic Research
          9712 Medical Center Dr, Rockville, MD 20850, USA
          Email: potato-array@tigr.org
          This clone can be obtained from the University of Arizona Genomics
          Institute. Orders can be made through URL:
          http://genome.arizona.edu/orders/
          Seq primer: T3.

FEATURES             source
    Location/Qualifiers
        1..754
            /organism="Solanum tuberosum"
            /mol_type="mRNA"
            /cultivar="Kennebec"
            /db_xref="taxon:4113"
            /clone="BPL117P18"
            /tissue_type="leaf"
            /dev_stage="6 week old"
            /lab_host="SOLR"
            /clone_lib="P. infestans-challenged potato leaf,
            incompatible reaction"
            /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
            XhoI; supplier: UC Berkeley, PGEC; sequencing: The
            Institute for Genomic Research. Whole plants were
            challenged with 450,000 sporangia/ml P. infestans isolate
            US-1 (US940501) in Biotron (Madison, Wisconsin). Leaf
            tissue was collected at 1, 2, 5, 12, and 24 hours
            post-challenge and frozen in liquid nitrogen immediately
            upon removal. Kennebec plants showed no signs of HR.
            Katahdin plants (susceptible to P. infestans US-1) were
            used as controls and showed infection. NOTE: We cannot
            exclude the possibility that this sequence is actually
            derived from Phytophthora rather than potato."

ORIGIN
Alignment Scores:
Pred. No.:      27.9      Length:      754
Score:          79.50      Matches:      29
Percent Similarity: 38.05%      Conservative: 14
Best Local Similarity: 25.66%      Mismatches: 33
Query Match:      16.91%      Indels:    37
DB:              12         Gaps:       4

US-09-917-376-4 (1-89) x BQ047524 (1-754)

QY      7  ValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln 26
      ::::

```

Db	152	GTGCTTATCATCAAGATCCAGGAAGT---CCGGTAGTAATCTAATATCTCCTGGTTCACTA	208
Qy	27	LeuValAsnThrGlySerSerSer-----	34
Db	209	GTATCAATTCGTGCACCTCTCACTTCCCTGGAAATGCCCTATATTAGTTTCGT	268
Qy	35	-----ValAspLeuSerThrValThrValArgTyrTrpPheThr	47
Db	269	AAGGGGGAGCCTCCGAAGTTTCTTGGTTATGAACATTTCTCCACTCGCAAAATGGGGTTCA	328
Qy	48	ArgaspGly-----GlySerSerThr	54
Db	329	AGGGTTGCTCAGATCATTTGACACCAAGTGCTGGGCTCTAGGCTAGGTTCTGGAAT	388
Qy	55	LeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleAargAlaSerPheGly	74
Db	399	CTGACCCCAATGTGGGATTTCAAGGCTAGGTTCTGGTACTGCTGACTCCAAATGGTGG	448
Qy	75	SerValAsnProAlaThrProThrAlaAspThrTyrLeu	87
Db	449	GAA-----CTCCTTCCGAGATAGTTACCTT	475
RESULT 3	CB293085	768 bp mRNA linear EST 28-FEB-2003	
LOCUS	UCRCS01_05cb12_g1	Washington Navel orange cold acclimated flavedo & albedo cDNA library Citrus sinensis cDNA clone UCRCS01_05cb12, mRNA	
DEFINITION		sequence.	
ACCESSION	CB293085		
VERSION	CB293085.1	GI:28618542	
KEYWORDS	EST.		
SOURCE	Citrus sinensis		
ORGANISM	Citrus sinensis		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA library was made, and 1 million primary lambda cDNA clones were in vivo excised to give a population of phluescript SK(-) phagemids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside (Fenton). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3530 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanmaker) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:	37.9	Length:	768
Pred. No.:	78.50	Matches:	20
Score:	57.69%	Conservative:	10
Percent Similarity:	38.46%	Mismatches:	15
Best Local Similarity:	16.70%	Indels:	7
Query Match:	14	Gaps:	2
DB:			

US-09-917-376-4 (1-89) x CB293085 (1-768)

Qy 3 GlyGlyValLysValGlnTyrLysAsnAsnAppSerAlaProGlyAspAsnGlnIleLys 22

Db 612 GCGGCGATCAAC-----AAGAACAAACAGCAGCAGCAATATAATAAAGTAAT 662

Qy 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42

Db 663 AATAATAGGACGCTGTCTACTAGTGTGCTAGGCTGTGTGCAAGTGGCGTACG--- 719

Qy 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54

Db 720 -----TTTCGGAAGCGGAGGAGCAGCAGT 746

RESULT 4

BO722991

LOCUS

DEFINITION

AGENCOURT\_8103733 Lupski sympathetic\_trunk Homo sapiens cDNA clone

IMAGE:6191052 5', mRNA sequence.

BO722991

BO722991.1

GI:21861888

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 906)

AUTHORS

NH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM13590 row: n column: 13

High quality sequence stop: 512.

Location/Qualifiers

1..906

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

FEATURES

source

```

/clone="IMAGE:6191052"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski_sympathetic trunk"
/notes="Vector: pCMV-SPOr6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTTCAGTCCGAGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 73.1 Length: 906
Score: 77.00 Matches: 16
Percent Similarity: 49.15% Conservative: 13
Best Local Similarity: 27.12% Mismatches: 26
Query Match: 16.38% Indels: 4
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x BQ722991 (1-906)

QY 7 ValGinTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuGln 26
DB 548 CTGCGCTATACCAATAAATCTCCAGGAGCCCGTGGAGTA-----ATGAAG 598
QY 27 LeuValAsnThrGlySerSerValAspSerValThrValThrValArgTyrTrpPhe 46
DB 599 GCTGTAAAGTGGATGAAGAATATGAGTTGGAAGCTCTTCCAGCACTACTGTAC 658
QY 47 ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
DB 659 TCCCGG--GGCGGATGCGCCCGCAGCTCTACACCTGCATCTGGGGCAGTGTGGA 712

RESULT 5
AA043716 376 bp mRNA linear EST 11-MAY-1997
LOCUS zk61f10.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:487339 3' similar to gb:L10240 BASIGIN PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION AA043716.1 GI:1521725
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 376)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.P., Chiapelli,B.,
Chisoso,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevisan,B., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
8889549
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 212.
Location/Qualifiers
1. .376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3761053"
/db_xref="taxon:9606"
/clone="IMAGE:487339"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/notes="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAAGAATTGCGGCGCGCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.: 28.5 Length: 376
Score: 76.00 Matches: 20
Percent Similarity: 43.42% Conservative: 13
Best Local Similarity: 26.32% Mismatches: 23
Query Match: 16.17% Indels: 20
DB: 9 Gaps: 3

US-09-917-376-4 (1-89) x AA043716 (1-376)

QY 17 GlyAspAsnGlnLeuLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36
DB 64 GGGGGTTCCACGTGAGCCCTGGT---GCTATAGANAGCCGCGCTCTCCAGGGGTGAG 120
QY 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
DB 121 CTGGGAGGGT---CTGCGGTCTGGAGTCCCGGCGATGGCGCAGTCCCGCAAC 177
QY 55 -----LeuValTyrAsn 58
DB 178 CCCTCCAGAGTCCCGCGGATGCACAGCAAGAGGGGCTTGGGAGTGACTTGAGGC 237
QY 59 CysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74
DB 238 TGTGACGGGTGCGCCCTCGGGTGTGGCAAGTGAAGTCTCTGTGGGC 285

RESULT 6
BE362891 606 bp mRNA linear EST 17-DEC-2001
LOCUS DGI_90_C11.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION BE362891
VERSION BE362891.2 GI:15724434
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 606)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
Pratt,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
On Jul 20, 2000 this sequence version replaced gi:9304448.
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology

```

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement.  
 Seq primer: PolyTMix  
 High quality sequence start: 15  
 High quality sequence stop: 594  
 POLYA=No.

#### FEATURES

Location/Qualifiers  
 1..606  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 55.4 Length: 606  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 10 Gaps: 4

US-09-917-376-4 (1-89) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 Db 208 ATAGCAATCAGCGGAACTCAATCAATCGGTCGGGAGGAGCGCTGTCTGCGCGGC 267  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 268 -----CGCGCGGAGCAGAACCTCGAACTTCCAAACGATCGGTGCAGAGGAA 315  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrValThrVal 61  
 Db 316 CCATGGTGGGATCTGCAGGAGGAGGTTTGACACATGATATATATATATATGATGA 375  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 376 GCTTGTGTGAAGTGTGTGACGTGTTCTTTGGAGACGGTGTCTGTCTGTGTGTCG 435  
 QY 79 AlaThrProThrAlaAsp 84  
 Db 436 AGCTTCGACGTCTGGTGTAT 453

#### RESULT 7

CD226529 641 bp mRNA linear EST 21-MAY-2003  
 LOCUS CCC1\_46\_E07.bl\_A007 Callus culture/cell suspension sorghum bicolor  
 DEFINITION CDNA clone CCC1\_46\_E07\_A007 3', mRNA sequence.  
 ACCESSION CD226529  
 VERSION CD226529.1 GI:30969963  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 641)  
 Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K., Eastman,A. and Pratt,L.H.  
 An EST database from Sorghum: callus culture and cell suspension  
 Unpublished (2003)

#### COMMENT

Other\_ESTs: CCC1\_46\_E07.g1\_A007  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1..641  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="RTx430"  
 /db\_xref="taxon:4558"  
 /clone="CCC1\_46\_E07\_A007"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Callus culture/cell suspension"  
 /note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from a mixture of polyA+ RNA from callus culture tissue and cells in suspension culture. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACCTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 59.9 Length: 641  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 14 Gaps: 4

US-09-917-376-4 (1-89) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeLysProGly 24  
 Db 149 ATAGCAATCAGCGGAACTCAATCAATCGGTCGGGAGGAGCGCTGTCTGCGCGGC 208  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 Db 209 -----CGCGCGGAGCAGAACCTCGAACTTCCAAACGATCGGTGCAGAGGAA 256  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrValThrVal 61  
 Db 257 CCATGGTGGGATCTGCAGGAGGAGGTTTGACACATGATATATATATATATGATGA 316  
 QY 62 AlaAlaMetGlyCysGlyAsnLe-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 317 GCTTGTGTGAAGTGTGTGACGTGTTCTTTGGAGACGGTGTCTGTCTGTGTGTCG 376  
 QY 79 AlaThrProThrAlaAsp 84  
 Db 377 AGCTTCGACGTCTGGTGTAT 394

RESULT 8  
 LOCUS BE362820  
 DEFINITION DGI\_89\_C02.g2\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA sequence.  
 ACCESSION BE362820  
 VERSION BE362820.1 GI:9304377

KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 657)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.  
 TITLE An EST database from Sorghum: dark-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 89  
 High quality sequence stop: 651  
 POLYA=No.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..657  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 62 Length: 657  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 10 Gaps: 4  
 US-09-917-376-4 (1-89) x BE362820 (1-657)  
 QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 DB 204 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGAGACCTGTGCTGCCCGCGGC 263  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
 DB 264 -----CGCGCGGGAGACAGAACCTGAACTTCAACACGATCGGTGCAAGGAAA 311  
 QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 DB 312 CCATGTGGGGATCTGCAGGAGAGAGGTTTGAGCAACATGAATATATATATGATGA 371  
 QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
 DB 372 GCTGTGTGAAGTGTGGTGTGCTGCTTTGAGACGCGTGTCTGTCTGTGTGTCGG 431  
 QY 79 AlaThrProThrAlaAsp 84  
 DB 432 AGCTTGCATCTGGTGAT 449  
 RESULT 9  
 CF487649  
 LOCUS  
 DEFINITION POLI\_45\_C08\_b1\_A002 Pollen Sorghum bicolor cDNA clone  
 POLI\_45\_C08\_A002 3', mRNA sequence.  
 CF487649  
 ACCESSION

CF487649.1 GI:34516518  
 EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 676)  
 AUTHORS Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J., Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and Pratt,L.H.  
 TITLE EST database from Sorghum: pollen  
 JOURNAL Unpublished (2003)  
 COMMENT Other ESTs: POLI\_45\_C08\_g1\_A002  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
 Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC)  
 POLYA=Yes.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..676  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="BTx623"  
 /db\_xref="taxon:4558"  
 /clone="POLI\_45\_C08\_A002"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Pollen"  
 /notes="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTx623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 64.6 Length: 676  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.17% Indels: 10  
 DB: 14 Gaps: 4  
 US-09-917-376-4 (1-89) x CF487649 (1-676)  
 QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 DB 214 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGAGACCTGTGCTGCCCGCGGC 273  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
 DB 274 -----CGCGCGGGAGACAGAACCTGAACTTCAACACGATCGGTGCAAGGAAA 321

```

QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 CCATGGTGGGGATCTGCAGGGAAGGAGTTTGAGCAACATGGAATATATATATGATGAA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 GCTTGTGTAAGTGTGGTGAGTGGTCTTTTGGAGACGGTGTCTGTGCTCTGTTGGTCCG 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 79 AlaThrProThrAlaAsp 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 AGCTTGCAGTCTGGTGAT 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
BG739392      691 bp  mRNA  linear  EST 15-MAY-2001
LOCUS  EMI_82_D05.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
DEFINITION  BG739392
ACCESSION  BG739392
VERSION    BG739392.1 GI:14089081
KEYWORDS   EST.
SOURCE     Sorghum bicolor (sorghum)
ORGANISM   Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE  1 (bases 1 to 691)
AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE     An EST database from Sorghum: developing embryos
JOURNAL   Unpublished (2000)
COMMENT   Contact: Cordonnier-Pratt MM
          Laboratory for Genomics and Bioinformatics
          The University of Georgia, Department of Plant Biology
          Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
          Tel: 706 542 1860
          Fax: 706 583 0210
          Email: mmpratt@uga.edu
          Sequences have been trimmed to exclude PolyA, vector and regions
          below Phred quality 16. The threshold for highest quality sequence
          is 20.
          Seq primer: PolyTMix
          High quality sequence start: 50
          High quality sequence stop: 691
          POLYA=No.

FEATURES             Location/Qualifiers
     source           1..691
                     /organism="Sorghum bicolor"
                     /mol_type="mRNA"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."

ORIGIN
Alignment Scores:
Pred. No.:          66.6      Length:          691
Score:              76.00     Matches:         23
Percent Similarity: 44.19%    Conservative:    15
Best Local Similarity: 26.74% Mismatches:         38
Query Match:        16.17%    Indels:           10
DB:                  12      Gaps:             4

US-09-917-376-4 (1-89) x BG739392 (1-691)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ATAGCAATCAGCCGAACATCAATGAATCCGGTCCGGGAGGAGCCTGTGCTGGCCCGGC 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg--- 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 325 -----CGGCCGGGACAGAACCTGAACTTCACAAACGATCGGTGCAAGAAA 372
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 CCATGGTGGGGATCTGCAGGGAAGGAGTTTGAGCAACATGGAATATATATATGATGAA 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 GCTTGTGTAAGTGTGGTGAGTGGTCTTTTGGAGACGGTGTCTGTGCTCTGTTGGTCCG 492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 79 AlaThrProThrAlaAsp 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 AGCTTGCAGTCTGGTGAT 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BG739392      724 bp  DNA  linear  GSS 16-JAN-2003
LOCUS  ig58e08.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
DEFINITION  ig58e08 5', genomic survey sequence.
ACCESSION  BG739392
VERSION    BG739392.1 GI:27766878
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 724)
AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Ballija,V., Dedhia,N.,
          Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
          Zutavern,T., McCombie,W.R. and Martienssen,R.A.
          Genomic shotgun sequences from Zea mays (methyl-filtered)
          Unpublished (2002)
          Contact: W. Richard McCombie
          Lita Annenberg Hazen Genome Sequencing Center
          Cold Spring Harbor Laboratory
          PO Box 100, Cold Spring Harbor, NY 11724, USA
          Tel: 516 367 8884
          Fax: 516 367 8874
          Email: mcombie@cshl.org
          Plate: ig58 row: e column: 08
          Seq primer: -21M13UnivRev
          Class: shotgun
          High quality sequence stop: 728.

FEATURES             Location/Qualifiers
     source           1..724
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="ig58e08"
                     /lab_host="DH5a"
                     /note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;
          The vector was digested with XbaI and one nucleotide was
          added by fill in in the recessive 3' end. The genomic DNA
          was rebligated, end repaired, adaptor ligated and size
          fractionated using sephadex. The resulting fragments were
          between 0.8 and 3 kb and were cloned into the vector (.x/y
          reads in M13mp19, .b/g reads in pUC19). The same ligation
          was transformed into DH5a."

ORIGIN
Alignment Scores:
Pred. No.:          71      Length:          724
Score:              76.00     Matches:         34
Percent Similarity: 35.28%    Conservative:     7
Best Local Similarity: 30.09% Mismatches:        29
Query Match:        16.17%    Indels:          43
DB:                  28      Gaps:             6

US-09-917-376-4 (1-89) x BG739392 (1-724)

```





QY 2 SerGlyGlyValIysValGlnTyrIysAsnAspSer-----AlaPro 16  
 DB 87 TCCGAGGCGTGCAGATCCCATACAGAGGAGCGAACAGCAACGCTGCGGAGCTGCGCC 146  
 QY 17 Gly-----AspAsnGlnIleIysProGlyLeuGlnLeuValAsn----- 29  
 DB 147 GCGCCCCAGCTGAGCAGCTCGACGGGAAGCCAGACTGATGCTTATCAACGCTGCGCCG 206  
 QY 30 Thr-GlySerSerValAsp----- 36  
 DB 207 ACGAGGAAGTTCGTCACCTCCCGGAGATCGTCGGAGCGGTCAGAAAGCTGGGTTCGAG 266  
 QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53  
 DB 267 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCGCGGACCGTCGAC 326  
 QY 53 rThrLeuValTyrAsnCysAspTTPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73  
 DB 327 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 357  
 QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 DB 358 -----CGGCGCTCACCACTTCTTCTTCC 381

RESULT 14  
 CG270741/c 934 bp DNA linear GSS 25-AUG-2003  
 LOCUS OGWIB33TH ZM 0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0583E18,  
 DEFINITION genomic survey sequence.

ACCESSION CG270741  
 VERSION CG270741.1 GI:34182882  
 KEYWORDS GSS.

SOURCE Zea mays  
 ORGANISM Zea mays

REFERENCE 1 (bases 1 to 934)  
 AUTHORS Whitelaw C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics  
 JOURNAL Unpublished (2002)  
 COMMENT Other GSSs: OGWIB33TV  
 Contact: Cathy Whitelaw

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org

Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..934  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0583E18"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 101 Length: 934  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 29 Indels: 43  
 DB: 29 Gaps: 6

US-09-917-376-4 (1-89) x CG270741 (1-934)

QY 2 SerGlyGlyValIysValGlnTyrIysAsnAspSer-----AlaPro 16  
 DB 804 TCCGAGGCGTGCAGATCCCATACAGAGGAGCGAACAGCAACGCTGCGGAGCTGCGCC 745  
 QY 17 Gly-----AspAsnGlnIleIysProGlyLeuGlnLeuValAsn----- 29  
 DB 744 GCGCCCCAGCTGAGCAGCTCGACGGGAAGCCAGACTGATGCTTATCAACGCTGCGCCG 685  
 QY 30 Thr-GlySerSerValAsp----- 36  
 DB 684 ACGAGGAAGTTCGTCACCTCCCGGAGATCGTCGGAGCGGTCAGAAAGCTGGGTTCGAG 625  
 QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53  
 DB 624 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCGCGGACCGTCGAC 565  
 QY 53 rThrLeuValTyrAsnCysAspTTPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73  
 DB 564 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 534  
 QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 DB 533 -----CGGCGCTCACCACTTCTTCTTCC 510

RESULT 15

CG270748

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

OTHER GSSs: OGWIB33TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES

Location/Qualifiers

1..1001

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0583E18"

/clone\_lib="ZM 0.7 1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 112 Length: 1001

Score: 76.00 Matches: 34

Percent Similarity: 36.28% Conservative: 7

Best Local Similarity: 30.09% Mismatches: 29

Query Match: 16.17% Indels: 43

DB: 29 Gaps: 6

US-09-917-376-4 (1-89) x CG270748 (1-1001)

```

QY 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 418 TCCGAGCGGTCCAGCATCCATACAAAGGAGGCGAACAGCAACGCTGCGCAGCTGCGCC 477
QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 478 GCGCGCCCGAGCTGAGCAGCTGAGCGGAGAGCCAGCAGCTGATGCTTATCAACCGTGGCCGC 537
QY 30 Thr-GlySerSerSerValAsp----- 36
Db 538 ACGAGGAAGTTCTGTTAACTTCCGGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTTCGAG 597
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 598 GTCATCCCGATCGAGCGCGGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 657
QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 658 TCGTGCG-----ACGTGCTCATGGCGCGCACGGGG----- 688
QY 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 689 -----CCGGCCTCACCACACTCTCTTCTCC 712

```

Search completed: May 12, 2004, 11:39:00  
Job time : 840.94 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 22.2257 Seconds  
(without alignments)  
2222.227 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspto.spool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QWMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOEM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376@cgn1\_116@runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTJUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	2029	3	US-09-136-574A-46
2	194.5	41.4	6416	3	US-09-136-574A-2
3	192.5	41.0	11707	3	US-09-136-574A-1
4	182.5	38.1	2977	2	US-07-862-588B-1
5	165	35.1	1438	4	US-09-339-159B-3
6	165	35.1	1482	3	US-09-198-956-9
7	165	35.1	1482	3	US-09-198-955A-11
8	165	35.1	1482	4	US-09-694-531-11
9	165	35.1	1482	4	US-09-670-141-9
10	165	35.1	1482	4	US-10-072-152-11
11	120.5	25.6	1624	2	US-07-862-588B-6
12	120.5	25.6	1775	2	US-07-862-588B-5

13	114	24.3	486	1	US-08-048-164A-1	Sequence 1, Appli
c 14	114	24.3	486	1	US-08-048-164A-3	Sequence 3, Appli
15	114	24.3	486	1	US-08-460-462-1	Sequence 1, Appli
c 16	114	24.3	486	1	US-08-460-462-3	Sequence 3, Appli
17	114	24.3	486	1	US-08-460-457-1	Sequence 1, Appli
c 18	114	24.3	486	1	US-08-460-457-3	Sequence 3, Appli
19	114	24.3	486	1	US-08-460-458-1	Sequence 1, Appli
c 20	114	24.3	486	1	US-08-460-458-3	Sequence 3, Appli
21	114	24.3	486	2	US-08-460-455-1	Sequence 1, Appli
c 22	114	24.3	486	2	US-08-460-455-3	Sequence 3, Appli
23	114	24.3	486	2	US-08-330-394A-1	Sequence 1, Appli
c 24	114	24.3	486	2	US-08-330-394A-3	Sequence 3, Appli
25	114	24.3	499	3	US-09-006-636-5	Sequence 5, Appli
26	114	24.3	499	3	US-09-006-632-5	Sequence 5, Appli
27	114	24.3	499	4	US-09-325-274-5	Sequence 21, Appl
28	113	24.0	1146	3	US-09-277-716-21	Sequence 21, Appl
29	113	24.0	1146	4	US-09-609-161B-21	Sequence 3, Appli
30	77	16.4	8107	4	US-09-335-586-3	Sequence 33, App
31	75.5	16.1	5714	4	US-09-620-312D-393	Sequence 21, Appl
32	72.5	15.4	20284	4	US-09-526-193A-21	Sequence 39, Appl
33	70	14.9	25165	4	US-09-453-702B-39	Sequence 20, Appl
34	67.5	14.4	2241	2	US-08-838-219B-20	Sequence 20, Appl
35	67.5	14.4	2241	3	US-09-233-336A-20	Sequence 20, Appl
36	67.5	14.4	2241	3	US-09-233-752A-20	Sequence 20, Appl
37	67.5	14.4	2241	3	US-09-402-036-20	Sequence 20, Appl
38	67.5	14.4	2241	4	US-09-904-226-20	Sequence 20, Appl
39	67.5	14.4	2370	2	US-08-838-219B-19	Sequence 19, Appl
40	67.5	14.4	2370	3	US-09-233-336A-19	Sequence 19, Appl
41	67.5	14.4	2370	3	US-09-233-752A-19	Sequence 19, Appl
42	67.5	14.4	2370	3	US-09-402-036-19	Sequence 19, Appl
43	67.5	14.4	2370	4	US-09-904-226-19	Sequence 19, Appl
44	67.5	14.4	118067	4	US-09-497-855A-32	Sequence 32, Appl
45	66	14.0	711	1	US-08-061-092A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.

```

;
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US0001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
;
US-09-136-574A-46
;
Alignment Scores:
Pred. No.: 1,06e-16 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.60% Indels: 1
DB: 3 Gaps: 1
;
US-09-917-376-4 (1-89) x US-09-136-574A-46 (1-2029)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAGAACATGAGCAAGTCGACGACGAGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ATAAGCCCGTGGTTAAGATAGTGAATGAGGCGAGCAGCAGTGTTCATCTTAGCAGGGTT 120
QY 41 ThrValArgTyrTipPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AGATAAGTACTGGTACACAGTGGATGGTGTGACACGCCACAGACTCGGTA---TGTGAC 177
QY 61 TPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCATTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
||| ||||| ||||| ||||| |||||
Db 238 AGTGGAGCGGATTATTACCTGGAG 261
;
RESULT 2
US-09-136-574A-2
; Sequence 2, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
;
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
;
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Alignment Scores:
Pred. No.: 7.26e-16 Length: 6416
Score: 194.50 Matches: 38
Percent Similarity: 61.18% Conservative: 14
Best Local Similarity: 44.71% Mismatches: 32
Query Match: 41.38% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-2 (1-6416)

QY 4 GlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2665 GGTGTGAGGTACTGTACAAAGCAATGAGACAGTGCAGACAGGTTCATTATAGCCG 2724
QY 24 GlyLeuGlnValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2725 TGGTTTAAAGATAGTGAATGGAGGCAGCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGA 2784
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAa 63
Db 2785 TACTGTATACACATGTGATGGTGACACAGCCACAGAGTGCAGTA---TGTGACTGGGCACAG 2841
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2842 ATAGGGGCAAGCAATGTGACATTCAATTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2901
QY 84 AspThrTyrLeuGln 88
Db 2902 GATTATTACCTGGAG 2916

RESULT 3
US-09-136-574A-1
; Sequence 1, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House

```

```
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 3 15e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservative: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 40.96% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-1 (1-11707)
QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTGTTGAAGTACTATACAAAGCAATGAGCAAGTGCAGCAGCAAGTCTCTATAAGCGG 4097
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGGTTTAAAGATAGTGAATGGAGCGACAGCAGTGTGATCTTAGCAGGGGTTAAGATAAGA 4157
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63
Db 4158 TACTGTACACAGTGTGATGGTGAAGCAGCAGCAGAGTGGCGTA---TGTGACTGGGCACAG 4214
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGGGCAAGCAATGTGCATTCATTTTGTGAAGCTGAGCAGCGGAGTGAGTGAGCGG 4274
QY 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.

; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
; US-07-862-588B-1

Alignment Scores:
Pred. No.: 1 05e-14 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.27% Conservative: 19
Best Local Similarity: 43.37% Mismatches: 25
Query Match: 38.83% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-1 (1-2977)
QY 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGACTGCATGACAGCGCGCGGATCAAAATGACGCGCACCAACAGATCAAGCGCTCTTC 203
QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAAACAACGCGTACTTCGCTGTTGATTTAAGCAGCGCTCAAAATCCGCTACTAC 263
QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGATGGTTCTCGCGCGGTGAACGGCTGG---ATCGACTGGCGGACGCTCGGC 320
QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GCAGCAACATTCAGATCTCGTTTGGC-----AACCTACTGCACGAATTCGATACG 374
QY 86 TyrLeuGln 88
Db 375 TACGTGGAG 383
```

```

; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-9

Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: Gaps: 2

US-09-917-376-4 (1-89) x US-09-198-956-9 (1-1482)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGTTGAATCTACACAGCAATCCTTCAGATACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTATACCGGAAGCAGTGCATTTGTTCCAAATC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACGACGACGACGACGACGACGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATATCGCAGTACCGCAGCTACACGGAATTTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAATAACGACGACACCTACCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-198-955A-11

```



```
Alignment Scores:
Pred. No.: 9,37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-198-955A-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTGAATTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAGGTACTAATACCGAAGCAGTGCAATGATTGTTGCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGCAGTAACGCGAGTACACGGAATTAATTCAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGCGACACCTACCTTGAA 1278

RESULT 8
US-09-694-531-11
; Sequence 11, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-694-531-11
Alignment Scores:
Pred. No.: 9,37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-694-531-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTGAATTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAGGTACTAATACCGAAGCAGTGCAATGATTGTTGCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGCAGTAACGCGAGTACACGGAATTAATTCAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGCGACACCTACCTTGAA 1278

RESULT 9
US-09-670-141-9
; Sequence 9, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-670-141-9
Alignment Scores:
Pred. No.: 9,37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-4 (1-89) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTGAATTCTACACAGCAATCCTTCAGATACTACTAATCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1111 ACATTGAGATATTATTATACAGTACGCGACGAGAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGCAGTAACGCGAGTACACGGAATTAATTCAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGCGACACCTACCTTGAA 1278
```

Db 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGACAGCTCAACGGAATTACTTCAAAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTACTCTTGAA 1278

## RESULT 10

US-10-072-152-11  
; Sequence 11, Application US/10072152  
; Patent No. 6677147  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schuelein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6677147el Pectate Lyases  
; FILE REFERENCE: 5378 200-US  
; CURRENT APPLICATION NUMBER: US/10/072,152  
; CURRENT FILING DATE: 2002-02-07  
; PRIOR APPLICATION NUMBER: US/09/198,955  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-10-072-152-11

## Alignment Scores:

Pred. No.: 9,37e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservative: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 4 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGGCAATTGGAAGTTGAATTTCAACAGCAATCTCTCAGATACTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 1111 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGACAGCTCAACGGAATTACTTCAAAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTACTCTTGAA 1278

## RESULT 11

US-07-862-588B-6  
; Sequence 6, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linea  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,588B  
; FILING DATE: 19920727  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 164/90  
; FILING DATE: 19-JAN-1990  
; PRIOR APPLICATION DATA: PCT/DK91/00013  
; FILING DATE: 18-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zelson, Steve T. / Lambiris, Elias J.  
; REGISTRATION NUMBER: 30,335 / 33,728  
; REFERENCE/DOCKET NUMBER: 3425.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1624 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus lautus  
; STRAIN: NCIMB 40250  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 30..1607  
; OTHER INFORMATION:  
US-07-862-588B-6

## Alignment Scores:

Pred. No.: 1.17e-06 Length: 1624  
Score: 120.50 Matches: 28  
Percent Similarity: 54.12% Conservative: 18  
Best Local Similarity: 32.94% Mismatches: 36  
Query Match: 25.64% Indels: 3

```

DB:      2      Caps:      2
US-09-917-376--4 (1-89) x US-07-862-588B-6 (1-1624)

Qy      2 SerGlyGlyValGlnTyrLysAenAenAspSerAlaProGlyAspAenGlnIle 21
Db      1323 ACGGGGAACCTTGTTGCTCAATACAAAGTTGGCGACACTAGCGCCACGGATAACCAANTG 1382
Qy      22 LysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrValThr 41
Db      1383 AAGCCTTCCTTTAAACATCAAGAACACCGGTACAAACCCCTGTTAACTGAGCGCCTCAAG 1442
Qy      42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAenCysAspTrp 61
Db      1443 CTTNNNNNNNNNNNNNAAAAGAC---GGACCTGCGGATATGAGCTCTCGATCGACTGG 1499
Qy      62 AlaAlaMetGlyCysGlyAenIleAaGAlaSerPheGlySerValAenProAlaThrPro 81
Db      1500 GCCAAATCGCGGAACGAATGTTCTGTGGCATTC-----GCTAACTTTACCGGGAGT 1553
Qy      82 ThrAlaAspThrTyr 86
Db      1554 AATACGGATACTTAC 1568

RESULT 12
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DX 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus

```

```

; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservatives: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 29.26% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-1 (1-486)
QY 2 SerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 7 ACATCATCAATGTCAGTTGAAATTTTACAACTCTAACAAATCAGCACAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 67 ACACCAATTAATCAAAATTAATACTACACATCTGACAGTGATTTAAATTTAAATGACGTAATA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 127 GTTAGATATTATACAAAGTGTGTTGATACAGGACAACTTCTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 184 GCTGTGCATTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

RESULT 14
US-08-048-164A-3/c
; Sequence 3, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
; US-08-048-164A-3

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservatives: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 29.26% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-3 (1-486)
QY 2 SerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 480 ACATCATCAATGTCAGTTGAAATTTTACAACTCTAACAAATCAGCACAACAACTCAATT 421
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 420 ACACCAATTAATCAAAATTAATACTACACATCTGACAGTGATTTAAATTTAAATGACGTAATA 361
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 360 GTTAGATATTATACAAAGTGTGTTGATACAGGACAACTTCTGG---TGTGACCAT 304
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 303 GCTGTGCATTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTC 244
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 243 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 196

RESULT 15
US-08-460-462-1
; Sequence 1, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-462-1

Alignment Scores:  
Pred. No.: 1.64e-06 Length: 486  
Score: 114.00 Matches: 29  
Percent Similarity: 52.58% Conservative: 22  
Best Local Similarity: 29.90% Mismatches: 34  
Query Match: 24.28% Indels: 12  
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-460-462-1 (1-486)

QY	2	SerGlyGlyValLysValClnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle	21
DB	7	ACATCATCAATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAACTCAATT	66
QY	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41
DB	67	ACACCAATAATCAAAATTACTTAACACTCTGACAGTGATTAAATTTAAATGACGTAAA	126
QY	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61
DB	127	GTTAGATATTATTACACAAAGTCATGGTACACAAAGCACAAACTTTCTGG--TGTGACCAT	183
QY	62	AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe	73
DB	184	GCTGGTGCATTATTAGGAAATAGCTATGTTGATAACACTAGCAAAAGTGACAGCAAACTTC	243
QY	74	-----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88
DB	244	GTTAAGAAACAGCAAGGCCA---ACATCAACCTATGATACATATGTTGAA	291

Search completed: May 12, 2004, 15:15:25  
Job time : 27.2257 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 136.654 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPTPTADTYLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376@cgn\_1\_809\_@runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

ALIGNMENTS

1	469	99.8	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.1	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	99.1	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.3	2869	13	US-09-917-376-2	Sequence 2, Appli
5	462	98.3	2869	15	US-10-155-400-2	Sequence 2, Appli
6	257	54.7	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.7	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	190	40.4	1527	16	US-10-369-493-46838	Sequence 46838, A
9	165	35.1	1438	13	US-10-372-054-3	Sequence 3, Appli
10	165	35.1	1482	13	US-10-655-433-11	Sequence 11, Appl
11	165	35.1	1482	14	US-10-072-152-11	Sequence 11, Appl
12	157	33.4	1314	9	US-09-974-300-684	Sequence 684, App
13	155.5	33.1	4059	10	US-09-784-554B-3	Sequence 1, Appli
14	150.5	32.0	4056	10	US-09-784-554B-3	Sequence 3, Appli
15	130.5	27.8	4452	13	US-10-282-122A-15299	Sequence 15299, A
16	114	24.3	768	15	US-10-261-446-5	Sequence 5, Appli
17	113	24.0	1146	10	US-09-808-898-21	Sequence 21, Appl
18	78.5	16.7	3147	13	US-10-282-122A-13716	Sequence 13716, A
19	76	16.2	96597	12	US-10-052-483-226	Sequence 226, App
20	75.5	16.1	1041	16	US-10-260-238-742	Sequence 742, App
21	75.5	16.1	2091	13	US-10-211-463-222	Sequence 222, App
22	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
23	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appl
24	75.5	16.1	4041	13	US-10-312-312-21	Sequence 21, Appl
25	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
26	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App
27	75.5	16.1	4914	13	US-10-116-802-38	Sequence 38, Appl
28	75.5	16.1	5714	9	US-09-747-835A-19	Sequence 19, Appl
29	75.5	16.1	5714	9	US-09-747-835A-60	Sequence 60, Appl
30	75.5	16.1	5714	13	US-10-312-312-19	Sequence 19, Appl
31	75.5	16.1	5714	13	US-10-312-312-60	Sequence 60, Appl
32	75.5	16.1	5714	15	US-10-037-270-393	Sequence 393, App
33	75.5	16.1	5714	16	US-10-117-722-393	Sequence 393, App
34	75.5	16.1	5773	9	US-09-747-835A-18	Sequence 18, Appl
35	75.5	16.1	5773	13	US-10-312-312-18	Sequence 18, Appl
36	75	16.0	484	13	US-10-424-599-54043	Sequence 54043, A
37	73	15.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
38	72.5	15.4	402	9	US-09-764-860-85	Sequence 85, Appl
39	72.5	15.4	402	15	US-10-074-095-85	Sequence 85, Appl
40	72.5	15.4	402	16	US-10-212-872-85	Sequence 85, Appl
41	72.5	15.4	9181	10	US-09-984-827-10	Sequence 10, Appl
42	72.5	15.4	20284	13	US-10-617-334-21	Sequence 21, Appl
43	72.5	15.4	20284	16	US-10-452-510-21	Sequence 21, Appl
44	72	15.3	2402	13	US-10-425-114-9462	Sequence 9462, Ap
45	72	15.3	48667	9	US-09-822-268A-3	Sequence 3, Appli

## Alignment Scores:

Pred. No.: 2,21e-56 Length: 2289  
 Score: 469.00 Matches: 88  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.79% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-378-2 (1-2289)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
 DB 1363 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACCG 1422  
 QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
 DB 1423 ATCAAAACCGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTGTTCGACGGTG 1482  
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1483 ACCGTGCGGTACTGGTTTACCCGGGATGGTGGTCTGTCGACACTGCTGTACAACTGTGAC 1542  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1543 TGGCGCGCGATGGGGTGGGAATATCCGGCCTCGTTCGGTGGTGAACCCGCGGACG 1602  
 QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1603 CCGACGGCGGACACCTACCTGCAG 1626

## RESULT 2

US-09-917-384-2

; Sequence 2, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 3687  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment of

US-09-917-384-2

; OTHER INFORMATION: Guxa

US-09-917-384-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
 Score: 466.00 Matches: 87  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.86% Mismatches: 0  
 Query Match: 99.15% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-384-2 (1-3687)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
 DB 1750 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACCG 1809  
 QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
 DB 1810 ATCAAAACCGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTGTTCGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1870 ACCGTGCGGTACTGGTTTACCCGGGATGGTGGTCTGTCGACACTGCTGTCAACTGTGAC 1929  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1930 TGGCGCGCGATGGGGTGGGAATATCCGGCCTCGTTCGGTGGTGAACCCGCGGACG 1989  
 QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 3

US-09-917-383-2

; Sequence 2, Application US/09917383  
 ; Publication No. US20030104522A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,383  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 3687  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment of

US-09-917-383-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
 Score: 466.00 Matches: 87  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.86% Mismatches: 0  
 Query Match: 99.15% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-383-2 (1-3687)

QY 1 ValSerGlyGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAenGln 20  
 DB 1750 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACCG 1809  
 QY 21 IleLysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrVal 40  
 DB 1810 ATCAAAACCGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTGTTCGACGGTG 1869  
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1870 ACCGTGCGGTACTGGTTTACCCGGGATGGTGGTCTGTCGACACTGCTGTCAACTGTGAC 1929  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1930 TGGCGCGCGATGGGGTGGGAATATCCGGCCTCGTTCGGTGGTGAACCCGCGGACG 1989  
 QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 4

US-09-917-376-2

; Sequence 2, Application US/09917376  
 ; Publication No. US20040038334A1



```
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGATAATGATTCGGCGCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCGGGTTTGCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTTGTCACGCGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGTTTACCCGGGATGGTGGCTCGTCGACACTGGTGTACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGTGTGGGAATATCCGCGCCTCGTTCGGCTCGGTGAACCCGCGCAGC 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACTCTGCAG 2868

RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:

; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGATAATGATTCGGCGCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCGGGTTTGCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTTGTCACGCGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGTTTACCCGGGATGGTGGCTCGTCGACACTGGTGTACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGTGTGGGAATATCCGCGCCTCGTTCGGCTCGGTGAACCCGCGCAGC 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACTCTGCAG 2868

RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2223)
US-10-156-761-550

Alignment Scores:
Pred. No.: 3,69e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
DB: 15 Gaps: 0
```

```
US-09-917-376-4 (1-89) x US-10-156-761-550 (1-2223)
QY 2 SerGlyValValysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1771 TCGGGCGGCTCAAGTCTCTTACAAGAACCAAGACTCTCGGCCACCGACCAACGCCATC 1830
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1831 CGGCCAGGCTTCGGATCGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTCCAG 1890
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
Db 1891 GCCCGCTACTACTTACCCGGGACAGCGGCTCGCCACCGTGAACCGCTGGTGGCGACTAC 1950
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1951 GCGGGCGTGGCTGTTCCAAAGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTCCAG 2010
QY 82 ThrAlaAspThrTyrLeuGln 88
Db 2011 GGAGCGGACGCTTACCTCGAA 2031

RESULT 7
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 2,72e-21 Length: 9025608
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
Gaps: 15

US-09-917-376-4 (1-89) x US-10-156-761-1 (1-9025608)
QY 2 SerGlyValValysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 706216 TCGGGCGGCTCAAGTCTCTTACAAGAACCAAGACTCTCGGCCACCGACCAACGCCATC 706275
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 706276 CGGCCAGGCTTCGGATCGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTCCAG 706335
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
Db 706336 GCGCGCTACTACTTACCCGGGACAGCGGCTCGCCACCGTGAACCGCTGGTGGCGACTAC 706395
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 706396 GCGGGCGTGGCTGTTCCAAAGTCAACCGCGAGCGGCTCCCTCGACCTGTCCAAAGGTCCAG 706455
QY 82 ThrAlaAspThrTyrLeuGln 88
Db 706456 GGAGCGGACGCTTACCTCGAA 706476

RESULT 8
US-10-369-493-46838
; Sequence 46838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46838
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-369-493-46838

Alignment Scores:
Pred. No.: 8.02e-17 Length: 1527
Score: 190.00 Matches: 35
Percent Similarity: 65.12% Conservative: 21
Best Local Similarity: 40.70% Mismatches: 26
Query Match: 40.43% Indels: 4
Gaps: 2

US-09-917-376-4 (1-89) x US-10-369-493-46838 (1-1527)
QY 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1084 GGTATTCTGTACGTACAGCAGGAGGGATGATGACAGACCAACCAATCGTCCG 1143
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1144 CAGCTTCAATAAAAAATAACGGCAATACCAACGGTTGATTTAAAAAGATGTCACTGCCGT 1203
QY 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62
Db 1204 TACTGGTATAAGCAAAAAACAAAGGCCAAAC-----TTTGACTGTGACTACGCG 1254
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 1255 CAGATTGGATCGGCAATGTGACACACAAGTTTGTGAGTTGTGATTAACCAACCAAGCAAGT 1314
QY 83 AlaAspThrTyrLeuGln 88
Db 1315 GCAGATACCTATCTGGAA 1332

RESULT 9
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
```

```

; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. I633
; US-10-372-054-3

Alignment Scores:
Pred. No.: 2,728-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x US-10-372-054-3 (1-1438)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGCAATTGAGGTGAATCTACACAGCAATCCTTCAGATACTACTCA 999
QY 21 IleLysProGlyLeuGlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTCAAGGTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 1059
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGCAGCAGAGAAGATCAGACCTTCTGG---TGTGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCATATATCGGCAGTACCGCAGCTACACCGGAATTACTTCAATGTAAAGGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAATGAGTTCTCTCAACAATAACGACGACACCTTACCTTGAA 1227

RESULT 10
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02

US-09-917-376-4 (1-89) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAGGTGAATCTACACAGCAATCCTTCAGATACTACTCA 1050
QY 21 IleLysProGlyLeuGlnValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTCAAGGTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGCAGCAGAGAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCATATATCGGCAGTACCGCAGCTACACCGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAATAACGACGACACCTTACCTTGAA 1278

RESULT 11
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217

```



```
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017-200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 1.298-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.02% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-3 (1-4056)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGATTGGTCTCCATGATCGCACGGGATACAAATGTGACGACAACTTGAACCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCAAATTTAAATAAGGTACATCTCCGTACCGATCAACGAGTTGAAATTCG 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 63
Db 3715 TACTACTACACGATCGCGGTGACCGTGAGCAG---ACATTCACTGCGACTATGCGGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3772 CTGAGCTGCTCGAAGCTGAATGGTAAAGCTGGTTAAATGGATAAAGCTGCAACCGGTGCT 3831
QY 84 AspThrTyrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 15
US-10-282-122A-15299
; Sequence 15299, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15299
; LENGTH: 4452
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15299

Alignment Scores:
Pred. No.: 1.04e-07 Length: 4452
Score: 130.50 Matches: 30
Percent Similarity: 55.79% Conservative: 23
Best Local Similarity: 31.58% Mismatches: 33
Query Match: 27.77% Indels: 9
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x US-10-282-122A-15299 (1-4452)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 100 AATTCGGGTGTTCAATACAAATTTGCTGATACAAATACTAGTACCAACCATGACTATT 159
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 160 GCTCTTAATTTAAATATCAAAATAATATCTGGAGCACCTTTAGATTTAACACTTTAAA 219
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 220 TTAAGATACTATTTTACAGCTGATGCTACTCAGGATGAAAATTTTGG---TGCACCAT 276
QY 62 AlaAlaMetGlyCysGly-----AsnIleArgAlaSerPhe 73
Db 277 GCTGGTATGCTTAATGGTTATACTACCAACAAATTAAGTAATGTAGTGGGTACTTTT 336
QY 74 GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 337 GTAGCTATGGTAATGCAACAGCTACTGCTGATCATTTACTTGAG 381

Search completed: May 12, 2004, 18:16:54
Job time : 817.654 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:10:32 ; Search time 54 Seconds  
(without alignments)  
465.680 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQVKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 10

Total number of hits satisfying chosen parameters: 12

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	98.9	88	7	ADD22925
2	88	98.9	89	7	ADD22924
3	88	98.9	154	6	ABP73020
4	88	98.9	762	6	ABP73022
5	83	93.3	150	6	ABP73025
6	83	93.3	1228	6	ABP73029
7	71	79.8	153	6	ABP71658
8	71	79.8	1121	6	ABP71656
9	36	40.4	88	6	ABP73018
10	36	40.4	89	6	ABP73017
11	36	40.4	957	6	ABP73015
12	36	40.4	957	7	ADD22921

ALIGNMENTS

RESULT 1  
ADD22925  
ID ADD22925 standard; protein; 88 AA.  
XX ADD22925;  
AC  
XX  
DT 15-JAN-2004 (first entry)  
XX

DE Acidothermus cellulolyticus avicelase AvIII CBD III #2.  
XX enzyme; AvIII; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-810853/76.  
DR  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 5; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase AvIII CBD III #2.  
XX  
SQ Sequence 88 AA;  
Query Match 98.9%; Score 88; DB 7; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.9e-81;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTDRDGSSTLVNCD 60  
DB 1 VSGGVKQVKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTDRDGSSTLVNCD 60  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
RESULT 2  
ADD22924  
ID ADD22924 standard; protein; 89 AA.  
XX  
AC ADD22924;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.  
KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 89

```
FT XX /label= Unknown
PN XX US2003108988-A1.
XX XX 12-JUN-2003.
XX XX 18-OCT-2002; 2002US-00155400.
XX XX 28-JUL-2001; 2001US-00917376.
XX XX (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
DR WPI; 2003-810853/76.
XX New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding Aviii and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX Claim 16; SEQ ID NO 4; 29pp; English.
XX The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable Aviii polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding Aviii. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase Aviii CBD III #1.
XX SQ Sequence 89 AA;
Query Match 98.9%; Score 88; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.9e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
Db 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 3
ABP73020
ID ABP73020 standard; peptide; 154 AA.
XX AC ABP73020;
XX XX 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the ManA carbohydrate binding domain type III.
XX ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannase.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012110-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023819.
XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannase.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-248182/24.
XX DR N-PSDB; ABZ77633.
XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
(MIDE ) MIDWEST RES INST.
Ding S, Adney WS, Vinzant TB, Himmel ME;
WPI; 2003-248182/24.
Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
cellulolyticus, useful for reducing hemicellulose in a starting material,
for processing of food, and as bulking agents in food stuffs.
Claim 8; Page 7; 46pp; English.
The present sequence represents a fragment of ManA, a thermostable
mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is
a member of the glycoside hydrolase family of enzymes. ManA is useful for
reducing hemicellulose in a starting material to simpler carbohydrate
units, and ultimately to sugars which are useful in the food, feed, paper
pulp, and biofuels industries. It is useful for the processing of food
and in food stuffs as bulking agents, and for the degradation of mannase.
ManA is also useful to raise polyclonal and monoclonal antibodies that
are useful in purifying ManA, or detecting ManA polypeptide expression,
and as well as reagent tools for characterizing the molecular actions of
CC ManA polypeptides
XX SQ Sequence 154 AA;
Query Match 98.9%; Score 88; DB 6; Length 154;
Best Local Similarity 100.0%; Pred. No. 3.2e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
Db 1 VSGGVKQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 4
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX AC ABP73022;
XX XX 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the ManA polypeptide.
XX ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannase.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012110-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023819.
XX PR 28-JUL-2001; 2001WO-US023819.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-248182/24.
XX DR N-PSDB; ABZ77633.
XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
```



XX PS Claim 3; Page 18-19; 46pp; English.

XX CC The present sequence represents Mana, a thermostable mannanase A

XX CC polypeptide derived from Acidothermus cellulolyticus. Mana is a member of

XX CC the glycoside hydrolase family of enzymes. Mana is useful for reducing

XX CC hemicellulose in a starting material to simpler carbohydrate units, and

XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and

XX CC biofuels industries. It is useful for the processing of food and in food

XX CC stuffs as bulking agents, and for the degradation of mannan. Mana is

XX CC also useful to raise polyclonal and monoclonal antibodies that are useful

XX CC in purifying Mana, or detecting Mana polypeptide expression, and as well

XX CC as reagent tools for characterizing the molecular actions of Mana

XX CC polypeptides

XX SQ Sequence 762 AA;

Query Match 98.9%; Score 88; DB 6; Length 762;

Best Local Similarity 100.0%; Pred. No. 1.4e-80;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGGSSTLVNCD 60

Db 455 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGGSSTLVNCD 514

Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542

RESULT 5

ABP73025

ID ABP73025 standard; peptide, 150 AA.

XX AC ABP73025;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the GuxA carbohydrate binding domain type III.

XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;

XX KW detergent; pulp processing; paper processing; feed processing; textile.

XX OS Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023817.

XX PR 28-JUL-2001; 2001WO-US023817.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus

XX cellulolyticus, useful for reducing cellulose in a starting material, and

XX for the conversion of biomass to biofuels and biofuel additives.

XX Claim 6; Page 7; 47pp; English.

XX ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a

XX thermostable cellulase, and is a member of the glycoside hydrolase family

XX of enzymes. GuxA is useful for reducing cellulose in a starting material

XX such as agricultural biomasses to sugars. This is useful in biofuel

XX production. GuxA is also useful in the conversion of biomass to biofuels

XX and biofuel additives, in detergents, pulp and paper processing, food and

XX feed processing, and in textile process. GuxA is also useful for raising

XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,

CC or detecting GuxA polypeptide expression, as well as reagent tools for

CC characterizing the molecular actions of GuxA polypeptides

XX SQ Sequence 150 AA;

Query Match 93.3%; Score 83; DB 6; Length 150;

Best Local Similarity 100.0%; Pred. No. 3.7e-76;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGGSSTLVNCDWAAMG 65

Db 6 KVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGGSSTLVNCDWAAMG 65

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88

Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 6

ABP73029

ID ABP73029 standard; protein, 1228 AA.

XX AC ABP73029;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the GuxA polypeptide.

XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;

XX KW detergent; pulp processing; paper processing; feed processing; textile.

XX OS Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023817.

XX PR 28-JUL-2001; 2001WO-US023817.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus

XX cellulolyticus, useful for reducing cellulose in a starting material, and

XX for the conversion of biomass to biofuels and biofuel additives.

XX Claim 3; Page 19; 47pp; English.

XX The present sequence represents a GuxA polypeptide. GuxA is thermostable

XX cellulase, and is a member of the glycoside hydrolase family of enzymes.

XX GuxA is useful for reducing cellulose in a starting material such as

XX agricultural biomass to sugars. This is useful in biofuel production.

XX GuxA is also useful in the conversion of biomass to biofuels and biofuel

XX additives, in detergents, pulp and paper processing, food and feed

XX processing, and in textile process. GuxA is also useful for raising

XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,

XX or detecting GuxA polypeptide expression, as well as reagent tools for

XX characterizing the molecular actions of GuxA polypeptides

XX SQ Sequence 1228 AA;

Query Match 93.3%; Score 83; DB 6; Length 1228;

Best Local Similarity 100.0%; Pred. No. 2.5e-75;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGGSSTLVNCDWAAMG 65

Db 589 KQYKNDSDAFQDNQIKPGLQVNTGSSVDLSVTYRYWFTDRGSSSTLYVNCDAAMG 648

QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
 |||||

Db 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 7

ABP71658

ID ABP71658 standard; protein; 153 AA.

AC ABP71658;

XX

DT 29-MAY-2003 (first entry)

XX

DE A. cellulolyticus Gux1 protein CBD\_III domain fragment.

XX

KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

KW biofuel; detergent; pulp; paper processing; feed processing; textile;

KW cellulase; enzyme.

XX

OS Acidothermus cellulolyticus.

XX

PN WO2003012095-A1.

XX

PD 13-FEB-2003.

XX

PF 28-JUL-2001; 2001WO-US023820.

XX

PR 28-JUL-2001; 2001WO-US023820.

XX

PA (MIDE ) MIDWEST RES INST.

XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

PI Lantz Mccarter S;

XX

DR WPI; 2003-300494/29.

XX

PT New thermal tolerant Gux1 peptide having specified amino acid sequence,

PT useful in the degradation of cellulose to biofuels.

XX

PS Claim 2; Page 7; 44pp; English.

XX

CC The invention relates to a thermal tolerant Gux1 peptide from A.

CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or

CC for conversion of biomass to biofuel additives. It is used in detergents,

CC pulp and paper processing, food and feed processing, and in textile

CC processing. It can also be used alone or in combination with other

CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-

CC scale processing of cellulose to sugar for use in biofuel production. The

CC present sequence represents a A. cellulolyticus Gux1 cellulase CBD\_III

CC domain fragment

XX

SQ Sequence 153 AA;

Query Match 79.8%; Score 71; DB 6; Length 153;

Best Local Similarity 100.0%; Pred. No. 6e-64;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 DNOIKPGLQVNTGSSVDLSVTYRYWFTDRGSSSTLYVNCDAAMCGNIRASFGSVN 77

Db 17 DNOIKPGLQVNTGSSVDLSVTYRYWFTDRGSSSTLYVNCDAAMCGNIRASFGSVN 76

QY 78 PATPTADTYLQ 88  
 |||||

Db 77 PATPTADTYLQ 87

RESULT 8

ABP71656

ID ABP71656 standard; protein; 1121 AA.

XX

AC ABP71656;

XX

DT 29-MAY-2003 (first entry)

XX

DE A. cellulolyticus Gux1 protein.

XX

KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

KW biofuel; detergent; pulp; paper processing; feed processing; textile;

KW cellulase; enzyme.

XX

OS Acidothermus cellulolyticus.

XX

FT Key Location/Qualifiers

FT Peptide 1..34

FT Protein /note= "potential signal peptide"

FT 35..1121

FT Misc-difference 228

FT /note= "mature protein"

FT /note= "encoded by CG"

XX

PN WO2003012095-A1.

XX

PD 13-FEB-2003.

XX

PF 28-JUL-2001; 2001WO-US023820.

XX

PR 28-JUL-2001; 2001WO-US023820.

XX

PA (MIDE ) MIDWEST RES INST.

XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

PI Lantz Mccarter S;

XX

DR WPI; 2003-300494/29.

DR N-PSDB; ABZ76162.

XX

PT New thermal tolerant Gux1 peptide having specified amino acid sequence,

PT useful in the degradation of cellulose to biofuels.

XX

PS Claim 3; Page 18-19; 44pp; English.

XX

CC The invention relates to a thermal tolerant Gux1 peptide from A.

CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or

CC for conversion of biomass to biofuel additives. It is used in detergents,

CC pulp and paper processing, food and feed processing, and in textile

CC processing. It can also be used alone or in combination with other

CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-

CC scale processing of cellulose to sugar for use in biofuel production. The

CC present sequence represents a A. cellulolyticus Gux1 cellulase

XX

SQ Sequence 1121 AA;

Query Match 79.8%; Score 71; DB 6; Length 1121;

Best Local Similarity 100.0%; Pred. No. 3.6e-63;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 DNOIKPGLQVNTGSSVDLSVTYRYWFTDRGSSSTLYVNCDAAMCGNIRASFGSVN 77

Db 51 DNOIKPGLQVNTGSSVDLSVTYRYWFTDRGSSSTLYVNCDAAMCGNIRASFGSVN 110

QY 78 PATPTADTYLQ 88  
 |||||

Db 111 PATPTADTYLQ 121

RESULT 9

XX	Acidothermus cellulolyticus.
OS	
XX	
XX	Key Location/Qualifiers
FH	Key
FT	Misc-difference 89
FT	/note= "unspecified residue"
XX	
PN	WO2003012090-A2.
XX	
XX	13-FEB-2003.
PD	
XX	
XX	28-JUL-2001; 2001WO-US023818.
PF	
XX	
XX	28-JUL-2001; 2001WO-US023818.
PR	
XX	(MIDE ) MIDWEST RES INST.
XX	
XX	Ding S, Adney WS, Vinzant TB, Himmel ME;
PI	
DR	WPI; 2003-248177/24.
XX	
XX	New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
PT	for degradation of cellulose or in generating anti-AvIII antibodies for
PT	purifying recombinant AvIII polypeptides from genetically engineered
PT	host cells.
PT	
XX	
XX	Claim 6; Page 8; 4app; English.
PS	
XX	The present sequence is derived from a thermostable avicelase, designated
CC	AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,
CC	and is a cellulase. AvIII is useful in the conversion of biomass to
CC	biofuels and biofuel additives. It may be useful in the production of
CC	detergents, pulp and paper processing, food and feed processing and in
CC	textile processes. The thermostable AvIII peptide is useful in the
CC	degradation of cellulose, and in generating specific anti-AvIII
CC	antibodies that are useful in purifying recombinant AvIII polypeptides
CC	from genetically engineered host cells, in detecting AvIII polypeptide
CC	expression, as well as a reagent tool for characterizing the molecular
CC	actions of the polypeptide. The AvIII polynucleotide is useful as a
CC	source of probes or primers in various diagnostic assays
XX	
XX	Sequence 89 AA;
SQ	
	Query Match 40.4%; Score 36; DB 6; Length 89;
	Best Local Similarity 100.0%; Pred. No. 1.3e-28;
	Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	28 VNTGSSVDLSTVTVRVWFTRDGGSLTVNCDWAA 63 
Db	28 VNTGSSVDLSTVTVRVWFTRDGGSLTVNCDWAA 63 
RESULT 11	
ABP73015	
ID	ABP73015 standard; protein; 957 AA.
XX	
AC	ABP73015;
XX	
DT	03-JUN-2003 (first entry)
XX	
XX	Amino acid sequence of the avicelase AvIII.
DE	
XX	
KW	Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW	detergent; pulp processing; paper processing; feed processing; textile;
KW	cellulose.
XX	
XX	Acidothermus cellulolyticus.
OS	
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 957
FT	/note= "unspecified residue encoded by N"
XX	
PN	WO2003012090-A2.

XX 13-FEB-2003.  
 PD 28-JUL-2001; 2001WO-US023818.  
 XX 28-JUL-2001; 2001WO-US023818.  
 XX (MIDE ) MIDWEST RES INST.  
 PA Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-248177/24.  
 DR N-PSDB; ABZ77632.  
 XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-AvIII antibodies for  
 PT purifying recombinant AvIII polypeptides from genetically engineered  
 PT host cells.  
 XX Claim 2; Page 20; 44pp; English.  
 XX The present sequence represents a thermostable avicelase polypeptide,  
 CC designated AvIII. AvIII is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. AvIII is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable AvIII peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC AvIII antibodies that are useful in purifying recombinant AvIII  
 CC polypeptides from genetically engineered host cells, in detecting AvIII  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The AvIII polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays  
 XX Sequence 957 AA;  
 SQ  
 Query Match 40.4%; Score 36; DB 6; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 28 VNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDAWAA 63  
 DB 896 VNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDAWAA 931  
 RESULT 12  
 ADD22921  
 ID ADD22921 standard; protein; 957 AA.  
 XX  
 AC ADD22921;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Acidothermus cellulolyticus avicelase AvIII.  
 XX enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 957  
 FT /label= Unknown  
 FT /note= "Encoded by N"  
 XX  
 PN US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX

PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-810853/76.  
 DR N-PSDB; ADD22922.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 16; SEQ ID NO 1; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.  
 XX Sequence 957 AA;  
 SQ  
 Query Match 40.4%; Score 36; DB 7; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 28 VNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDAWAA 63  
 DB 896 VNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDAWAA 931  
 Search completed: May 11, 2004, 12:26:52  
 Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:23:57 ; Search time 21 Seconds  
(without alignments)  
407.669 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKXNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR 78.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: May 11, 2004, 12:28:43  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:11:27 ; Search time 15 Seconds  
(without alignments)  
308.950 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
-----					

No matches found

Search completed: May 11, 2004, 12:27:19  
Job time : 16 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:23:22 ; Search time 39 Seconds  
(without alignments)  
720.029 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKYQKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

No matches found

Search completed: May 11, 2004, 12:28:10  
Job time : 39 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:25:52 ; Search time 23 Seconds  
(without alignments)  
199.770 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKVKYKNDSPAGDQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	†
No.	Score	Match Length DB ID Description
-----		

No matches found

Search completed: May 11, 2004, 12:29:19  
Job time : 23 secs



Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	88	98.9	88	12	US-09-917-376-5	Sequence 5, Appli
2	88	98.9	88	14	US-10-155-400-5	Sequence 5, Appli
3	88	98.9	89	12	US-09-917-376-4	Sequence 4, Appli
4	88	98.9	89	14	US-10-155-400-4	Sequence 4, Appli
5	88	98.9	154	10	US-09-917-378-4	Sequence 4, Appli
6	88	98.9	762	10	US-09-917-378-1	Sequence 1, Appli
7	83	93.3	150	10	US-09-917-384-5	Sequence 5, Appli
8	83	93.3	150	10	US-09-917-383-5	Sequence 5, Appli
9	83	93.3	1043	10	US-09-917-384-6	Sequence 6, Appli
10	83	93.3	1043	10	US-09-917-383-6	Sequence 6, Appli
11	83	93.3	1228	10	US-09-917-383-1	Sequence 1, Appli
12	83	93.3	1228	10	US-09-917-383-1	Sequence 1, Appli
13	36	40.4	957	12	US-09-917-376-1	Sequence 1, Appli
14	36	40.4	957	14	US-10-155-400-1	Sequence 1, Appli

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 3

US-09-917-376-4  
; Sequence 4, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-4

Query Match 98.9%; Score 88; DB 12; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 4

US-10-155-400-4  
; Sequence 4, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-4

Query Match 98.9%; Score 88; DB 14; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 5

US-09-917-378-4  
; Sequence 4, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate  
; OTHER INFORMATION: binding domain  
US-09-917-378-4

Query Match 98.9%; Score 88; DB 10; Length 154;  
Best Local Similarity 100.0%; Pred. No. 2.7e-78;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60  
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 6

US-09-917-378-1  
; Sequence 1, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus

```
US-09-917-378-1
Query Match      98.9%; Score 88; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 1.1e-77;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
Db 455 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 514

Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-5

Query Match      93.3%; Score 83; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.1e-73;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 6 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5

Query Match      93.3%; Score 83; DB 10; Length 1043;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 482 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 541

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match      93.3%; Score 83; DB 10; Length 1043;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 65
Db 482 KQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAAMG 541

Qy 66 CGNIRASFGSVNPATPTADTYLQ 88
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6
```



; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 40.4%; Score 36; DB 14; Length 957;  
Best Local Similarity 100.0%; Pred.No. 1.3e-26;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	28	VNTGSSVDLSVTYRYWFTRDGSGSTLVNCDWAA	63
Db	896	VNTGSSVDLSVTYRYWFTRDGSGSTLVNCDWAA	931

Search completed: May 11, 2004, 12:33:59  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:26:57 ; Search time 182 Seconds  
(without alignments)  
477.301 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 89

Sequence: 1 VSGGVKVOYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 6019581 seqs, 976053577 residues

Word size : 10

Total number of hits satisfying chosen parameters: 16

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents\_AA\_Main.\*

- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097A\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US097B\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US099A\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/2/paa/US099B\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*
- 28: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*
- 29: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*
- 30: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*
- 31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*
- 32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*
- 33: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

RESULT 1

US-09-917-376-5

; Sequence 5, Application US/09917376

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; FILE OF INVENTION: CELLULOYTICUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

; LENGTH: 88

; TYPE: PRT

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; OTHER INFORMATION: Carbohydrate binding domain

US-09-917-376-5

Query Match Similarity 98.9%; Score 88; DB 24; Length 88;

Best Local Similarity 100.0%; Pred. No. 8e-81; Mismatches 0; Indels 0; Gaps 0;

Matches 88; Conservative 0;

Qy 1 VSGGVKVOYKNNDSAPGDNQIKPGLQVNTGSSVDLSITVTVYVWTRDGGSTLYNCD 60

Db 1 VSGGVKVOYKNNDSAPGDNQIKPGLQVNTGSSVDLSITVTVYVWTRDGGSTLYNCD 60

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2

US-10-155-400-5

; Sequence 5, Application US/10155400

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS

; FILE OF INVENTION: CELLULOYTICUS

; FILE REFERENCE: NREL 01-36A

; CURRENT APPLICATION NUMBER: US/10/155,400

; CURRENT FILING DATE: 2002-10-22

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

```
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match      98.9%; Score 88; DB 27; Length 88;
Best Local Similarity 100.0%; Pred. No. 8e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      98.9%; Score 88; DB 24; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.1e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 4
US-10-155-400-4
; Sequence 4, Application US/10155400
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      98.9%; Score 88; DB 24; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.1e-81;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match      98.9%; Score 88; DB 24; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.3e-80;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYVNC 60
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
```

; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-1

Query Match 98.9%; Score 88; DB 24; Length 762;  
Best Local Similarity 100.0%; Pred. No. 5.8e-80;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCD 60  
Db 455 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCD 514  
Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7  
US-09-917-383-5  
; Sequence 5, Application US/09917383  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-5

Query Match 93.3%; Score 83; DB 24; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 6 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 8  
US-09-917-384-5  
; Sequence 5, Application US/09917384  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-5

Query Match 93.3%; Score 83; DB 24; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 6 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 66 CGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
US-09-917-383-6  
; Sequence 6, Application US/09917383  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-6

Query Match 93.3%; Score 83; DB 24; Length 1043;  
Best Local Similarity 100.0%; Pred. No. 9.4e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 6 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 65  
Db 482 KVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVNCDWAAMG 541  
Qy 66 CGNIRASFGSVNPATPTADTYLQ 88  
Db 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 10  
US-09-917-384-6  
; Sequence 6, Application US/09917384  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.



; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-6

Query Match 93.3%; Score 83; DB 24; Length 1043;  
Best Local Similarity 100.0%; Pred. No. 9.4e-75;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65  
DB 482 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 541  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 542 CGNIRASFGSVNPATPTADTYLQ 564

RESULT 11  
US-09-917-383-1  
; Sequence 1, Application US/09917383  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-1

Query Match 93.3%; Score 83; DB 24; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.1e-74;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65  
DB 589 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 648  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 12  
US-09-917-384-1  
; Sequence 1, Application US/09917384

; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-1

Query Match 93.3%; Score 83; DB 24; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.1e-74;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 65  
DB 589 KVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMG 648  
QY 66 CGNIRASFGSVNPATPTADTYLQ 88  
DB 649 CGNIRASFGSVNPATPTADTYLQ 671

RESULT 13  
US-09-917-384A-4  
; Sequence 4, Application US/09917384A  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: MCCARTER, SUZANNE  
; APPLICANT: HIMMEL, MICHAEL E.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: VINZANT, TODD B.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: NREL 01-38  
; CURRENT APPLICATION NUMBER: US/09/917,384A  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-384A-4

Query Match 79.8%; Score 71; DB 24; Length 153;  
Best Local Similarity 100.0%; Pred. No. 2.5e-63;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCCNIRASFGSVN 77  
DB 17 DNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDDWAAMGCCNIRASFGSVN 76  
QY 78 PATPTADTYLQ 88  
DB 77 PATPTADTYLQ 87

RESULT 14  
US-09-917-384A-1

```
; Sequence 1, Application US/09917384A
; GENERAL INFORMATION:
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: DING, SHI-YOU
; APPLICANT: MCCARTER, SUZANNE
; APPLICANT: HIMMEL, MICHAEL E.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: VINZANT, TODD B.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-38
; CURRENT APPLICATION NUMBER: US/09/917,384A
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1121
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; US-09-917-384A-1

Query Match      79.8%; Score 71; DB 24; Length 1121;
Best Local Similarity 100.0%; Pred. No. 1.6e-62;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 DNQIKPGLQVNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAAAGCGNIRASFGSVN 77
Db 51 DNQIKPGLQVNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAAAGCGNIRASFGSVN 110

Qy 78 PATPTADTYLQ 88
Db 111 PATPTADTYLQ 121

RESULT 15
US-09-917-376-1
; Sequence 1, Application US/09917376
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-09-917-376-1

Query Match      40.4%; Score 36; DB 24; Length 957;
Best Local Similarity 100.0%; Pred. No. 4.9e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 VNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAA 63
Db 896 VNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAA 931

RESULT 16
US-10-155-400-1
; Sequence 1, Application US/10155400
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
```

```
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-10-155-400-1

Query Match      40.4%; Score 36; DB 27; Length 957;
Best Local Similarity 100.0%; Pred. No. 4.9e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 VNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAA 63
Db 896 VNTGSSVDLSTVTTRYWFTRDGSSSTLVYNCDAWAA 931

Search completed: May 11, 2004, 12:32:34
Job time : 183 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:27:22 ; Search time 18 Seconds  
(without alignments)  
169.758 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKRNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 234425 seqs, 34333021 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents\_AA New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	----------	-------------

No matches found

Search completed: May 11, 2004, 12:33:04  
Job time : 18 secs



```
Db 1483 ACGTGGCGGTACTGTTTACCCGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGTAC 1542
Qy 61 TtpAlaAlaMetGlyCyseGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGGGGGGATGGGTGGGATATATCCGCGCTCGTTCGGTCTGTAACCCGGCGAGC 1602
Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
TITLE Thermal tolerant cellulase from Acidothermus cellulolyticus
JOURNAL Patent: WO 0301209-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1. 3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of Guxa"

ORIGIN
Alignment Scores:
Pred. No.: 3,236-78 Length: 3687
Score: 83.00 Matches: 83
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 93.26% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700036 (1-3687)
Qy 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 1765 AAGTGCAGTACAGAACAAATGATTCGGCGCGGTGTATACAGATCAACCGGGTCTC 1824
Qy 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 1825 CAGTTGGTGAATACCGGGTCTGTCGGTGGATTGTTCGACGGTACGGTGGTCTGG 1884
Qy 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 1885 TTCACCGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGACTGGCGCGGATGGGG 1944
Qy 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 1945 TGTGGGAATATCCGGCCCTCGTTCGGTCTGGTGAACCCGGCGACCGCGGACACC 2004
Qy 86 TyrLeuGln 88
Db 2005 TACCTGCAG 2013

RESULT 3
AX700025
LOCUS AX700025 3365 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION AX700025
VERSION AX700025.1 GI:29536018
KEYWORDS Acidothermus cellulolyticus
SOURCE Acidothermus cellulolyticus
ORGANISM Acidothermus cellulolyticus
```

```
REFERENCE 1
AUTHORS Adney, W.S., Ding, S.Y., Vinzant, T.B., Himmel, M.E., Decker, S.R. and Lantz, McCarter, S.
TITLE Thermal tolerant exoglucanase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012095-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1. 3365
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 1,826-65 Length: 3365
Score: 71.00 Matches: 71
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.78% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700025 (1-3365)
Qy 18 AspAsnGlnIleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeu 37
Db 151 GACAACCCAGATCAACACCGGGTCTCCAGTTGGTGAATACACCGGGTCTCGTGGTGGATTG 210
Qy 38 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 57
Db 211 TCGACGGTGCACGGTGGTCTGCTTCCACCGGGATGGTGGTCTGTCGACACTGGTGTAC 270
Qy 58 AsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 77
Db 271 AACTGTGACTGGCGCGGATGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGTGGAAC 330
Qy 78 ProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 331 CCGGGCGACCGCGCGGACACCTACCTGCAG 363

RESULT 4
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS Acidothermus cellulolyticus
ORGANISM Acidothermus cellulolyticus
SOURCE Acidothermus cellulolyticus
REFERENCE 1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant avicelase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012090-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1. 2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 3,096-28 Length: 2869
Score: 36.00 Matches: 36
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.45% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x AX700058 (1-2869)
```

Qy	28	VallAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThr	47
Db	2686	GTGAATACCGGTCGTCGTCGGATTGTTCGACGGTGACGGTCCGGTACTGGTTCACC	2745
Qy	48	ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2746	CGGGATGGTGGCTCGTCGACACTGGTGACAACTGTGACTGGGCGGCG	2793

Search completed: May 12, 2004, 16:05:36  
 Job time : 2810 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: May 12, 2004, 11:39:06 ; Search time 322 Seconds  
(without alignments)  
1174.191 Million cell updates/sec  
Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 30

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+ p2n.spool -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09917376/runat.11052004.114645.28650/app\_query.fasta.1.263  
-DB=N Geneseq 29Jan04 -QMT=fastcap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=1000 -DOCALLN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn.1\_470@runat.11052004.114645.28650 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq 29Jan04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	88	98.9	2289	7	Abz77633 Nucleotid
2	83	93.3	3687	7	Abz77634 Nucleotid
3	71	79.8	3365	7	Abz776162 A. cellul
4	36	40.4	2869	7	Abz77632 Nucleotid
5	36	40.4	2869	9	ADD22922 Acidothe

ALIGNMENTS

RESULT 1  
ABZ77633 standard; DNA; 2289 BP.  
XX ABZ77633;  
AC ABZ77633;  
XX 03-JUN-2003 (first entry)  
DT Nucleotide sequence of the Mana polypeptide.  
DE Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
KW food; feed; paper pulp; biofuel; mannanase; gene; ss.  
XX Acidotherrnus cellulolyticus.  
XX Key Location/Qualifiers  
FT CDS 1..2289  
FT /\*tag= a  
FT /product= "Mana"  
XX WO2003012110-A1.  
XX 13-FEB-2003.  
XX 28-JUL-2001; 2001WO-US023819.  
XX 28-JUL-2001; 2001WO-US023819.  
XX (MIDE ) MIDWEST RES INST.  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248182/24.  
XX P-FSDB; ABP73022.  
XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus  
PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
PT for processing of food, and as bulking agents in food stuffs.  
XX Example 1; Page 23; 46pp; English.  
XX The present sequence encodes Mana, a thermostable mannanase A polypeptide  
CC derived from Acidothermus cellulolyticus. Mana is a member of the  
CC glycoside hydrolase family of enzymes. Mana is useful for reducing  
CC hemicellulose in a starting material to simpler carbohydrate units, and  
CC ultimately to sugars which are useful in the food, feed, paper pulp, and  
CC biofuels industries. It is useful for the processing of food and in food  
CC stuffs as bulking agents, and for the degradation of mannanase. Mana is  
CC also useful to raise polyclonal and monoclonal antibodies that are useful  
CC in purifying Mana, or detecting Mana polypeptide expression, and as well  
CC as reagent tools for characterizing the molecular actions of Mana  
CC polypeptides  
XX Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 2.32e-77 Length: 2289  
Score: 88.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.88% Indels: 0  
DB: 7 Gaps: 0  
US-09-917-376-4 (1-89) x ABZ77633 (1-2289)  
Qy 1 ValserGlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 1363 GGTGCGGTGGGGTGAAGTGCAGTACACACATGATTCGGCGCGGGTGATACCAG 1422  
Qy 21 lleylsProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1423 ATCAAAACCGGCTCTCCAGTTGGTGAATACGGGGTCTGCTCGGTGGATTGTGACGGTG 1482

QY 41 ThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1483 ACGGTGCGGTACTGTTCCACCGGGATGTTGGTCTGTCGACACTGTTACAACTGTGAC 1542

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1543 TGGGCGGCGATGGGTGTGGGAATATCCGGCTCGTTCCGCTCGGTGAACCGCGGACG 1602

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1603 CCGACGGCGGACACTACTGTCAG 1626

RESULT 2  
 ABZ77634  
 ID ABZ77634 standard; DNA; 3687 BP.  
 XX  
 AC ABZ77634;  
 XX  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Nucleotide sequence of the Guxa polypeptide.  
 XX  
 KW Guxa; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW gene; ss.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..3687  
 FT /\*tag= a  
 FT /product= "Guxa"  
 XX  
 PN WO2003012109-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023817.  
 XX  
 PR 28-JUL-2001; 2001WO-US023817.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
 XX  
 DR WPI; 2003-239526/23.  
 DR P-PSDB; ABP73029.  
 XX  
 PT Novel thermal tolerant Guxa polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.  
 XX  
 PS Example 1; Page 23-24; 47pp; English.  
 XX  
 CC The present sequence encodes a Guxa polypeptide. Guxa is thermostable  
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
 CC Guxa is useful for reducing cellulose in a starting material such as  
 CC agricultural biomass to sugars. This is useful in biofuel production.  
 CC Guxa is also useful in the conversion of biomass to biofuels and biofuel  
 CC additives, in detergents, pulp and paper processing, food and feed  
 CC processing, and in textile processes. Guxa is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying Guxa,  
 CC or detecting Guxa polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of Guxa polypeptides  
 XX  
 SQ Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3,238-72 Length: 3687  
 Score: 83.00 Matches: 83  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 93.26% Indels: 0  
 DB: 7 Gaps: 0  
 US-09-917-376-4 (1-89) x ABZ77634 (1-3687)

QY 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25  
 DB 1765 AAGGTGCGAGTCAAGAACAAATGATTCGGCGCGGGTGATAACAGATCAAACCGGGTCTC 1824

QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTyr 45  
 DB 1825 CAGTTGGTGAATACCGGGTCTGCTCGTGGATTCTGTCACGGTGCAGTTCGGTACTGG 1884

QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65  
 DB 1885 TTCACCGGGATCGTGGTCTGTCGACACTGGTGTACAACTGTCTACTGGCGCGATGGG 1944

QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85  
 DB 1945 TGTGGGAATATCCGGCTCGTTCGGTCCGTGAACCCGCGACCGCGGACACC 2004

QY 86 TyrLeuGln 88  
 DB 2005 TACCTGCAG 2013

RESULT 3  
 ABZ76162  
 ID ABZ76162 standard; DNA; 3365 BP.  
 XX  
 AC ABZ76162;  
 XX  
 DT 29-MAY-2003 (first entry)  
 XX  
 DE A. cellulolyticus Gux1 protein encoding DNA.  
 XX  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; gene; ds.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..3365  
 FT /\*tag= a  
 FT /product= "Gux1"  
 FT /transl\_except= (pos: 682..683, aa: Pro)  
 FT /note= "this codon has an apparent one nucleotide  
 FT basepair deletion which alters the reading frame"  
 XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023820.  
 XX  
 PR 28-JUL-2001; 2001WO-US023820.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz Mccarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 DR P-PSDB; ABP71656.  
 XX  
 PT New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Disclosure; Page 22-23; 44pp; English.  
 XX  
 CC The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate



CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
 CC DNA

XX Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,29e-60 Length: 3365  
 Score: 71.00 Matches: 71  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 79.78% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ76162 (1-3365)

QY 18 AsplandnlllelvsProGlyLeuValenThrGlySerSerValAspLeu 37  
 DB 151 GACAAACAGATCAACCGGGTCTCCAGTTGGTGAATACCGGGTGGTCTGGTGGATTG 210  
 QY 38 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 57  
 DB 211 TCGACGGTGAAGGCGCGGTACTGTGTTACCCGGGATGGTGGTCTGCACACTGGTGAC 270  
 QY 58 AsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 77  
 DB 271 AACTGTGACTGGGGGGCGATGGGGTGTGGGATATCCCGCGCTGTTCCGGTCTGGTGAAC 330

QY 78 ProIaThrProThrAlaAspThrTyrLeuGln 88

DB 331 CCGGCGAGCGCGGCGGACACCTACCTGCAG 363

## RESULT 4

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX AC ABZ77632;

XX DT 03-JUN-2003 (first entry)

XX DE Nucleotide sequence of the avicelase AvIII.

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose; gene; ss.

XX OS Acidothermus cellulolyticus.

XX PH Key Location/Qualifiers

XX CDS 1..2869

XX FT /\*tag= a

XX FT /partial

XX FT /product= "AvIII"

XX FT /transl\_except= (pos:2869,aa:Xaa)

XX FT /note= "Xaa is an unspecified residue"

XX PN WO2003012090-A2.

XX XX 13-FEB-2003.

XX XX 28-JUL-2001; 2001WO-US023818.

XX XX 28-JUL-2001; 2001WO-US023818.

XX XX (MIDE ) MIDWEST RES INST.

XX

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-248177/24.

XX P-PSDB; ABP73015.

XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-AvIII antibodies for  
 PT purifying recombinant AvIII polypeptides from genetically engineered  
 PT host cells.

XX Claim 3; Page 24; 44pp; English.

XX The present sequence encodes a thermostable avicelase polypeptide,  
 CC designated AvIII. AvIII is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. AvIII is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable AvIII peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC AvIII antibodies that are useful in purifying recombinant AvIII  
 CC polypeptides from genetically engineered host cells, in detecting AvIII  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The AvIII polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays

XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

## Alignment Scores:

Pred. No.: 9,28e-26 Length: 2869  
 Score: 36.00 Matches: 36  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 40.45% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-4 (1-89) x ABZ77632 (1-2869)

QY 28 ValAenThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThr 47

DB 2686 GTGAATACCGGGTCTGTCGTGGTGGATTGTGACGGTGCAGTGGTACTGTTCCACC 2745

QY 48 ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63

DB 2746 CGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGACTGGGCGCG 2793

## RESULT 5

ADD22922

ID ADD22922 standard; DNA; 2869 BP.

XX AC ADD22922;

XX DT 15-JAN-2004 (first entry)

XX DE Acidothermus cellulolyticus avicelase AvIII DNA.

XX KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;  
 KW glycoside hydrolase; avicelase; ds; gene.

XX OS Acidothermus cellulolyticus.

XX PH Key Location/Qualifiers

XX CDS 1..2868

XX FT /\*tag= a

XX FT /product= "AvIII"

XX PN US2003108988-A1.

XX XX 12-JUN-2003.

XX XX 18-OCT-2002; 2002US-00155400.

XX XX 28-JUL-2001; 2001US-00917376.

XX

PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-810853/76.  
DR P-PSDB; ADD22921.  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
XX Claim 17; SEQ ID NO 2; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents DNA encoding the Acidothermus cellulolyticus  
CC avicelase AvIII.  
XX  
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:  
Pred. No.:                      9.28e-26                      Length:                      2869  
Score:                          36.00                      Matches:                      36  
Percent Similarity:            100.00%                    Conservative:                  0  
Best Local Similarity:        100.00%                    Mismatches:                   0  
Query Match:                   40.45%                    Indels:                        0  
DB:                              9                              Gaps:                        0

US-09-917-376-4 (1-89) x ADD22922 (1-2869)

QY    28 ValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThr 47  
Db    2686 GTGATACCGGGTCGTGCGGTGATTGTTCGACGGTGACGGTGGGTACTGGTTACCC 2745

QY    48 ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db    2746 CGGGATGGTGGCTGCTCGACACTGGGTGTACAACTGTGACTGGGGCGCG 2793

Search completed: May 12, 2004, 15:18:40  
Job time : 324 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 15:15:37 ; Search time 1965 Seconds  
(without alignments)  
1352.536 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 89

Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frames\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114646\_28669/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376@cgn\_1\_1\_3437@runat\_11052004\_114646\_28669 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPOP=60 -DELEXT=7  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_hic:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_hic:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: em\_gss\_hum:\*\*
- 18: em\_gss\_inv:\*\*
- 19: em\_gss\_pln:\*\*
- 20: em\_gss\_vrt:\*\*
- 21: em\_gss\_fun:\*\*
- 22: em\_gss\_mam:\*\*
- 23: em\_gss\_mus:\*\*
- 24: em\_gss\_pro:\*\*
- 25: em\_gss\_rod:\*\*
- 26: em\_gss\_phg:\*\*
- 27: em\_gss\_vrl:\*\*
- 28: gb\_gss1:\*\*

29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
------------	-------	-------	--------------	-------	-------------

No matches found

Search completed: May 12, 2004, 16:38:30  
Job time : 1968 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 15:18:47 ; Search time 67 Seconds  
(without alignments)  
737.174 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 89  
Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_11052004\_114646\_28682/app.query.fasta\_1.263  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olip2n.rni -MINWATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=1000 -DOCALIGN=200 -THR SCORE=quality -THR MIN=30 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1.69 @runat\_11052004\_114646\_28682 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: May 12, 2004, 16:39:50  
Job time : 69 secs

Result	Query	Score	Length	DB	ID	Description
No.						

DB Z 748 C G G G A T G G T G G C T C G T C G A C A C T G G T G T A C A A C T G G G C G G C G Z 753

RESULT 5  
US-10-155-400-2  
; Sequence 2, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-10-155-400-2

Alignment Scores:  
Pred. No.: 1,08e-27 Length: 2869  
Score: 36.00 Matches: 36  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 40.45% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-10-155-400-2 (1-2869)

QY	28	ValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThr	47
Db	2686	GTGAATACCGGGTCGTCGGTGGATTGTGCGACGGTGACGGTGGGACTGGTTAC	2745
QY	48	ArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2746	CGGGATGGTGGCTCGTGACACACTGGGTACAACTGTGACTGGGCGCG	2793

Search completed: May 12, 2004, 17:28:07  
Job time : 349 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 14.4613 Seconds  
(without alignment)  
1738.900 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	7 ADD22925	Add22925 Acidother
2	469	99.8	89	7 ADD22924	Add22924 Acidother
3	469	99.8	154	6 ABP73020	Abp73020 Amino aci
4	469	99.8	762	6 ABP73022	Abp73022 Amino aci
5	466	99.1	150	6 ABP73025	Abp73025 Amino aci
6	466	99.1	1228	6 ABP73029	Abp73029 Amino aci
7	462	98.3	88	6 ABP73018	Abp73018 Amino aci
8	462	98.3	89	6 ABP73017	Abp73017 Amino aci
9	462	98.3	957	7 ABP73015	Abp73015 Amino aci
10	462	98.3	957	7 ADD22321	Add22321 Acidother
11	436	92.8	153	6 ABP71658	Abp71658 A. cellul
12	436	92.8	1121	6 ABP71656	Abp71656 A. cellul
13	231.5	49.3	782	2 AAR15625	Aar15625 Cellulase
14	195.5	41.6	616	2 AAV13494	AAV13494 Truncated
15	195.5	41.6	616	2 AAE16325	AAE16325 Active ce
16	194.5	41.4	1751	2 AAV13493	AAV13493 Truncated
17	194.5	41.4	1751	5 AAE16324	AAE16324 Active ce
18	191.5	40.7	1426	2 AAV13492	AAV13492 Truncated
19	191.5	40.7	1426	5 AAE16323	AAE16323 Active ce
20	188	40.0	477	5 AAO22444	AAO22444 Protein e
21	188	40.0	496	5 AAO22443	AAO22443 Protein e
22	188	40.0	677	5 AAU98063	AAU98063 Bacillus
23	186	39.6	499	2 AAR42122	Aar42122 NK-1 cell
24	172	36.6	700	2 AAR13227	Aar13227 Novel end
25	165.5	35.2	551	2 AAW18790	AAW18790 Corrected

#### ALIGNMENTS

##### RESULT 1

ADD22925

ID ADD22925 standard; protein; 88 AA.

XX AC ADD22925;

XX XX 15-JAN-2004 (first entry)

XX DE Acidothermus cellulolyticus avicelase Aviii CBD III #2.

XX XX enzyme; Aviii; cellulose reduction; agricultural biomass;

KW municipal solid waste; glycoside hydrolase; avicelase.

XX OS Acidothermus cellulolyticus.

XX PN US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel MB;

DR WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for

PT detection of a polynucleotide encoding Aviii and for reducing cellulose

PT in a starting material, e.g. municipal solid waste.

XX Claim 16; SEQ ID NO 5; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a  
thermostable Aviii polypeptide. The polynucleotide is useful for  
detection of a polynucleotide encoding Aviii. The polynucleotide is  
useful for reducing cellulose in a starting material which involves  
administering to the starting material, e.g. agricultural biomass or  
municipal solid waste, a polypeptide molecule of the polynucleotide. The  
method further comprises administering a second polypeptide molecule  
chosen from the glycoside hydrolase family of proteins. The present  
sequence represents the amino acid sequence of Acidothermus  
cellulolyticus avicelase Aviii CBD III #2.

Aar95080 Cellulose  
Aay54123 A manana  
Aay28850 Pectate 1  
Aay43218 Pectate 1  
Aaw15238 Scaffold 1  
Aaw43108 C. thermo  
Aag63963 Amino aci  
Aag63963 Amino aci  
Aab23559 Protein e  
Aaw01503 60 kD end  
Aar13229 Endogluc  
Aar13229 Endogluc  
Aar63634 Cellulose  
Aaw90077 C. cellul  
Aae05745 Clostridi  
Aab81128 C17E2 Osp  
Aau97869 E. coli c  
Aae05749 Chimeric  
Aae05747 Clostridi  
Aae05748 Clostridi  
Aae05746 Clostridi

26 165 35.1 167 2 AAR95080  
27 165 35.1 476 3 AAY54123  
28 165 35.1 493 2 AAY28850  
29 165 35.1 493 2 AAY43218  
30 165 35.1 531 2 AAW15238  
31 165 35.1 1853 2 AAW43108  
32 155.5 33.1 1352 4 AAG63962  
33 150.5 32.0 1350 4 AAG63963  
34 130.5 27.8 1483 6 AAB23559  
35 120.5 25.6 531 2 AAW01503  
36 120.5 25.6 532 2 AAR13229  
37 114 24.3 162 2 AAR63634  
38 114 24.3 162 2 AAW90077  
39 114 24.3 163 4 AAE05745  
40 114 24.3 256 4 AAB81128  
41 114 24.3 256 5 AAU97869  
42 114 24.3 328 4 AAE05749  
43 114 24.3 341 4 AAE05747  
44 114 24.3 428 4 AAE05748  
45 113 24.0 190 4 AAE05746



```

XX SQ Sequence 88 AA;
Query Match 99.8%; Score 469; DB 7; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.9e-46;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60
Db 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2
ADD22924
ID ADD22924 standard; protein; 89 AA.
XX AC ADD22924;
XX DT 15-JAN-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.
XX KW enzyme; AvIII; cellulose reduction; agricultural biomass;
XX KW municipal solid waste; glycoside hydrolase; avicelase.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
FT Misc-difference 89 /label= Unknown
XX FT
XX PN US2003108988-A1.
XX PD 12-JUN-2003.
XX PF 18-OCT-2002; 2002US-00155400.
XX PR 28-JUL-2001; 2001US-00917376.
XX PA (DING/) DING S.
XX PA (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-810853/76.
XX DR New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX PS Claim 16; SEQ ID NO 4; 29pp; English.
XX CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase AvIII CBD III #1.
XX SQ Sequence 89 AA;
Query Match 99.8%; Score 469; DB 7; Length 89;
Best Local Similarity 100.0%; Pred. No. 4e-46;

XX SQ Sequence 88 AA;
Query Match 99.8%; Score 469; DB 6; Length 154;
Best Local Similarity 100.0%; Pred. No. 7.8e-46;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60
Db 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
ABP73020
ID ABP73020 standard; peptide; 154 AA.
XX AC ABP73020;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the ManA carbohydrate binding domain type III.
XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX KW food; feed; paper pulp; biofuel; mannanase.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012110-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023819.
XX PR 28-JUL-2001; 2001WO-US023819.
XX PA (WIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248182/24.
XX DR Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX PS Claim 8; Page 7; 46pp; English.
XX CC The present sequence represents a fragment of ManA, a thermostable
CC mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is
CC a member of the glycoside hydrolase family of enzymes. ManA is useful for
CC reducing hemicellulose in a starting material to simpler carbohydrate
CC units, and ultimately to sugars which are useful in the food, feed, paper
CC pulp, and biofuels industries. It is useful for the processing of food
CC and in food stuffs as bulking agents, and for the degradation of mannanase.
CC ManA is also useful to raise polyclonal and monoclonal antibodies that
CC are useful in purifying ManA, or detecting ManA polypeptide expression,
CC and as well as reagent tools for characterizing the molecular actions of
CC ManA polypeptides
XX SQ Sequence 154 AA;
Query Match 99.8%; Score 469; DB 6; Length 154;
Best Local Similarity 100.0%; Pred. No. 7.8e-46;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60
Db 1 VSGGVKQVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWFTRDGSSSTLVYNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

```

```

RESULT 4
ABP73022 ID ABP73022 standard; protein; 762 AA.
XX AC ABP73022;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the ManA polypeptide.
XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX KM food; feed; paper pulp; biofuel; mannanase.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012110-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023819.
XX PR 28-JUL-2001; 2001WO-US023819.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-248182/24.
XX DR N-PSDB; ABZ77633.
XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
XX PT cellulolyticus, useful for reducing hemicellulose in a starting material,
XX PT for processing of food, and as bulking agents in food stuffs.
XX PS Claim 3; Page 18-19; 46pp; English.
XX CC The present sequence represents ManA, a thermostable mannanase A
XX CC polypeptide derived from Acidothermus cellulolyticus. ManA is a member of
XX CC the glycoside hydrolase family of enzymes. ManA is useful for reducing
XX CC hemicellulose in a starting material to simpler carbohydrate units, and
XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and
XX CC biofuels industries. It is useful for the processing of food and in food
XX CC stuffs as bulking agents, and for the degradation of mannanase. ManA is
XX CC also useful to raise polyclonal and monoclonal antibodies that are useful
XX CC in purifying ManA, or detecting ManA polypeptide expression, and as well
XX CC as reagent tools for characterizing the molecular actions of ManA
XX CC polypeptides
XX SQ Sequence 762 AA;
Query Match 99.8%; Score 469; DB 6; Length 762;
Best Local Similarity 100.0%; Pred. No. 5.7e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVYNC 60
Db 455 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVYNC 514
QY 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMGCNIRASFGSVNPATPTADTYLQ 542
RESULT 5
ABP73025 ID ABP73025 standard; peptide; 150 AA.
XX AC ABP73025;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the GuxA carbohydrate binding domain type III.

```

```

KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012109-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023817.
XX PR 28-JUL-2001; 2001WO-US023817.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX DR WPI; 2003-239526/23.
XX PT Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX PT cellulolyticus, useful for reducing cellulose in a starting material, and
XX PT for the conversion of biomass to biofuels and biofuel additives.
XX PS Claim 6; Page 7; 47pp; English.
XX CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a
XX CC thermostable cellulase, and is a member of the glycoside hydrolase family
XX CC of enzymes. GuxA is useful for reducing cellulose in a starting material
XX CC such as agricultural biomass to sugars. This is useful in biofuel
XX CC production. GuxA is also useful in the conversion of biomass to biofuels
XX CC and biofuel additives, in detergents, pulp and paper processing, food and
XX CC feed processing, and in textile process. GuxA is also useful for raising
XX CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX CC or detecting GuxA polypeptide expression, as well as reagent tools for
XX CC characterizing the molecular actions of GuxA polypeptides
XX SQ Sequence 150 AA;
Query Match 99.1%; Score 466; DB 6; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.7e-45;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVYNC 60
Db 1 VSGGLKVOYKNNDSAPGDNQIKFGLQLVNTGSSVDLSVTIVRYWFTRDGSSSTLVYNC 60
QY 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
RESULT 6
ABP73029 ID ABP73029 standard; protein; 1228 AA.
XX AC ABP73029;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the GuxA polypeptide.
XX KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
XX KM detergent; pulp processing; paper processing; feed processing; textile.
XX OS Acidothermus cellulolyticus.
XX PN WO2003012109-A1.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023817.
XX PR 28-JUL-2001; 2001WO-US023817.

```

```

XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX PI WPI; 2003-239526/23.
XX DR N-PSDB; AB277634.
XX DR Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX PT cellulolyticus, useful for reducing cellulose in a starting material, and
XX PT for the conversion of biomass to biofuels and biofuel additives.
XX XX
XX PS Claim 3; Page 19; 47pp; English.
XX CC The present sequence represents a GuxA polypeptide. GuxA is thermostable
XX CC cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX CC GuxA is useful for reducing cellulose in a starting material such as
XX CC agricultural biomass to sugars. This is useful in biofuel production.
XX CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX CC additives, in detergents, pulp and paper processing, food and feed
XX CC processing, and in textile processes. GuxA is also useful for raising
XX CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX CC or detecting GuxA polypeptide expression, as well as reagent tools for
XX CC characterizing the molecular actions of GuxA polypeptides
XX SQ Sequence 1228 AA;
      Query Match 99.1%; Score 466; DB 6; Length 1228;
      Best Local Similarity 98.9%; Pred. No. 2.3e-44;
      Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNC 60
DB 584 VSGGLKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNC 643
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671
      RESULT 7
      ABP73018
      ID ABP73018 standard; peptide; 88 AA.
      AC ABP73018;
      DT 03-JUN-2003 (first entry)
      DE Amino acid sequence of AviiiI carbohydrate binding domain type II.
      KW Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
      KW detergent; pulp processing; paper processing; feed processing; textile;
      KW cellulose.
      OS Acidothermus cellulolyticus.
      XX
      FH Key Location/Qualifiers
      FT Misc-difference 89
      XX /note= "unspecified residue"
      PN WO2003012090-A2.
      XX
      PD 13-FEB-2003.
      XX
      PE 28-JUL-2001; 2001WO-US023818.
      XX
      PR 28-JUL-2001; 2001WO-US023818.
      XX
      PA (MIDE ) MIDWEST RES INST.
      XX
      PI Ding S, Adney WS, Vinzant TB, Himmel ME;
      XX WPI; 2003-248177/24.
      XX New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
      PT for degradation of cellulose or in generating anti-AviiiI antibodies for
      PT purifying recombinant AviiiI polypeptides from genetically engineered
      XX host cells.
      PS Claim 6; Page 8; 44pp; English.
      XX

```

---

```

PT XX host cells.
PS Claim 6; Page 8; 44pp; English.
XX CC The present sequence is derived from a thermostable avicelase, designated
XX CC AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes,
XX CC and is a cellulase. AviiiI is useful in the conversion of biomass to
XX CC biofuels and biofuel additives. It may be useful in the production of
XX CC detergents, pulp and paper processing, food and feed processing and in
XX CC textile processes. The thermostable AviiiI peptide is useful in the
XX CC degradation of cellulose, and in generating specific anti-AviiiI
XX CC antibodies that are useful in purifying recombinant AviiiI polypeptides
XX CC from genetically engineered host cells, in detecting AviiiI polypeptide
XX CC expression, as well as a reagent tool for characterizing the molecular
XX CC actions of the polypeptide. The AviiiI polynucleotide is useful as a
XX CC source of probes or primers in various diagnostic assays
XX SQ Sequence 88 AA;
      Query Match 98.3%; Score 462; DB 6; Length 88;
      Best Local Similarity 97.7%; Pred. No. 2.5e-45;
      Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNC 60
DB 1 VSGGVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
      RESULT 8
      ABP73017
      ID ABP73017 standard; peptide; 89 AA.
      AC ABP73017;
      DT 03-JUN-2003 (first entry)
      DE Amino acid sequence of AviiiI carbohydrate binding domain type III.
      KW Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
      KW detergent; pulp processing; paper processing; feed processing; textile;
      KW cellulose.
      OS Acidothermus cellulolyticus.
      XX
      FH Key Location/Qualifiers
      FT Misc-difference 89
      XX /note= "unspecified residue"
      PN WO2003012090-A2.
      XX
      PD 13-FEB-2003.
      XX
      PE 28-JUL-2001; 2001WO-US023818.
      XX
      PR 28-JUL-2001; 2001WO-US023818.
      XX
      PA (MIDE ) MIDWEST RES INST.
      XX
      PI Ding S, Adney WS, Vinzant TB, Himmel ME;
      XX WPI; 2003-248177/24.
      XX New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
      PT for degradation of cellulose or in generating anti-AviiiI antibodies for
      PT purifying recombinant AviiiI polypeptides from genetically engineered
      XX host cells.
      PS Claim 6; Page 8; 44pp; English.
      XX

```

CC The present sequence is derived from a thermostable avicelase, designated  
CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. Aviii is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable Aviii peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-Aviii  
CC antibodies that are useful in purifying recombinant Aviii polypeptides  
CC from genetically engineered host cells, in detecting Aviii polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The Aviii polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX  
SQ

Query Match 98.3%; Score 462; DB 6; Length 89;  
Best Local Similarity 97.7%; Pred. No. 2.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
DB 1 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAIGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
ABP73015  
ID ABP73015 standard; protein; 957 AA.  
XX  
AC ABP73015;  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Amino acid sequence of the avicelase Aviii.  
XX  
KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 957 /note= "unspecified residue encoded by N"  
FT  
XX WO2003012090-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 28-JUL-2001; 2001WO-US023818.  
XX  
PR 28-JUL-2001; 2001WO-US023818.  
XX  
PA (MIDE ) MIDWEST RES INST.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-248177/24.  
DR N-PSDB; AB277632.  
XX  
PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-Aviii antibodies for  
PT purifying recombinant Aviii polypeptides from genetically engineered  
PT host cells.  
XX  
PS Claim 2; Page 20; 44pp; English.  
XX  
CC The present sequence represents a thermostable avicelase polypeptide,  
CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of

CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable Aviii peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC Aviii antibodies that are useful in purifying recombinant Aviii  
CC polypeptides from genetically engineered host cells, in detecting Aviii  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
CC as a source of probes or primers in various diagnostic assays  
XX  
SQ

Query Match 98.3%; Score 462; DB 6; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60  
DB 869 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 928  
  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 10  
ADD22921  
ID ADD22921 standard; protein; 957 AA.  
XX  
AC ADD22921;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii.  
XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 957 /label= Unknown  
FT /note= "Encoded by N"  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-810853/76.  
DR N-PSDB; ADD22922.  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 1; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable Aviii polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.  
 XX  
 SQ Sequence 957 AA;

Query Match 98.3%; Score 462; DB 7; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLYNCD 60  
 DB 869 VSGGVKQVKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLYNCD 928  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 929 WAAMCGNIRASFGSVNPATPTADTYLQ 956

RESULT 11  
 ABP71658  
 ID ABP71658 standard; protein; 153 AA.  
 AC ABP71658;  
 XX  
 XX 29-MAY-2003 (first entry)  
 DE A. cellulolyticus Guxl protein CBD\_III domain fragment.  
 XX  
 XX Guxl; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 XX Acidothermus cellulolyticus.  
 OS  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023820.  
 XX  
 PR 28-JUL-2001; 2001WO-US023820.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 XX  
 XX New thermal tolerant Guxl peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Claim 2; Page 7; 44pp; English.  
 XX  
 CC The invention relates to a thermal tolerant Guxl peptide from A.  
 CC cellulolyticus. The Guxl exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Guxl cellulase CBD\_III  
 domain fragment  
 XX  
 SQ Sequence 153 AA;

Query Match 92.8%; Score 436; DB 6; Length 153;  
 Best Local Similarity 96.4%; Pred. No. 4.9e-42;  
 Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 VKVOYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLYNCDWAAM 64  
 DB 4 LKAQYKNNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLYNCDWAAM 63  
 QY 65 GCGNIRASFGSVNPATPTADTYLQ 88  
 DB 64 GCGNIRASFGSVNPATPTADTYLQ 87

RESULT 12  
 ABP71656  
 ID ABP71656 standard; protein; 1121 AA.  
 AC ABP71656;  
 XX  
 XX 29-MAY-2003 (first entry)  
 DE A. cellulolyticus Guxl protein.  
 XX  
 XX Guxl; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.  
 XX  
 XX Acidothermus cellulolyticus.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /note= "potential signal peptide"  
 FT Protein 35..1121  
 FT /note= "mature protein"  
 FT Misc-difference 228  
 FT /note= "encoded by CG"  
 XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023820.  
 XX  
 PR 28-JUL-2001; 2001WO-US023820.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 XX  
 DR WPI; 2003-300494/29.  
 DR N-PSDB; ABZ76162.  
 XX  
 XX New thermal tolerant Guxl peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Claim 3; Page 18-19; 44pp; English.  
 XX  
 CC The invention relates to a thermal tolerant Guxl peptide from A.  
 CC cellulolyticus. The Guxl exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Guxl cellulase  
 XX  
 SQ Sequence 1121 AA;

Query Match 92.8%; Score 436; DB 6; Length 1121;  
Best Local Similarity 96.4%; Pred. No. 5.7e-41;  
Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 VKVQYKNDSPGNDQIKPGQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNCDAAM 64  
DB 38 LKAQYKNDSPGNDQIKPGQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNCDAAM 97

QY 65 GCGNIRASFGSVNPATPTADTYLQ 88  
DB 98 GCGNIRASFGSVNPATPTADTYLQ 121

RESULT 13  
AAR15625  
ID AAR15625 standard; protein; 782 AA.  
XX  
AC AAR15625;  
XX  
XX 25-MAR-2003 (revised)  
DT 17-MAR-1992 (first entry)  
XX  
XX Cellulase AE-1.  
DE  
XX Detergents; pharmaceuticals; deinking; carboxymethylcellulose.  
KW Aeromonas strain no. 212.  
XX  
OS  
XX JP03251174-A.  
PN  
XX  
PD 08-NOV-1991.  
XX  
XX 28-FEB-1990; 90JP-00045465.  
PF  
XX 28-FEB-1990; 90JP-00045465.  
PR  
XX (OJIP ) OJI PAPER CO.  
PA  
XX WPI; 1991-373412/51.  
DR  
XX N-PSDB; AAQ15178.  
XX  
XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
PT pH when carboxymethylcellulose is used as substrate.  
XX  
XX Claim 2; Fig 3; 8pp; Japanese.  
PS  
XX The sequence was deduced from the gene which was sequenced from plasmid, PAEC 1, prepd. by ligating chromosomal DNA contg. the gene (obtd. from Aeromonas) into pUC18. The protein has mol. wt. of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-methylcellulose is the substrate. CC  
XX The N-terminal sequence: GHADT- has been confirmed by Edman degradation. CC  
XX The gene can be used to produce recombinant enzyme which is used for the CC effective utilis- ation of biomass resources and the mfr. of CC pharmaceuticals and foodstuffs, and also for the detergent and deinking CC of waste paper. (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 782 AA;

Query Match 49.3%; Score 231.5; DB 2; Length 782;  
Best Local Similarity 48.3%; Pred. No. 1.3e-17;  
Matches 42; Conservative 16; Mismatches 28; Indels 1; Gaps 1;

QY 2 SGGVKQYKNDSPGNDQIKPGQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNCDAAM 61  
DB 631 SGDLAVQYKTDNADNQPKPHFNVKNGAAAPLSLSLYFTAD-GNDQLQVNCDAAM 689

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 690 AAMGCGNIRASFGSVNPATPTADTYLQ 716

RESULT 14

AAY13494  
ID AAY13494 standard; protein; 616 AA.  
XX  
AC AAY13494;  
XX  
XX 30-JUL-1999 (first entry)  
DT  
XX Truncated cellulase Cel E3/B5.  
DE  
XX  
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
KW cotton-containing fabric; stonewashing.  
XX  
OS Unidentified.  
XX  
XX EP921188-A2.  
PN  
XX  
PD 09-JUN-1999.  
XX  
XX 15-SEP-1998; 98EP-00810919.  
PF  
XX 19-SEP-1997; 97US-00932571.  
PR  
XX (CLRN ) CLARIANT FINANCE BVI LTD.  
PA  
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
PI Morgan H, Williams DP;  
XX  
XX WPI; 1999-315403/27.  
DR  
XX N-PSDB; AAX55660.  
XX  
XX New truncated cellulase proteins, useful in detergents and for producing  
PT 'stonewashed' denim.  
XX  
XX Claim 7; Page 42-43; 65pp; English.

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic materials including cotton-containing fabrics. They are especially useful for preventing redeposition of colorant during stonewashing and for processing of textiles where cellulose breakdown is required. The new truncated enzymes show reduced redeposition of dye compared to using non-truncated cellulase compositions

SQ Sequence 616 AA;

Query Match 41.6%; Score 195.5; DB 2; Length 616;  
Best Local Similarity 43.2%; Pred. No. 1.3e-13;  
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVKQYKNDSPGNDQIKPGQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNCDAAM 60  
DB 1 MGSQGVKQYKNDSPGNDQIKPGQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNCDAAM 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 60 WAQIGASNTVFNPKLSSGVSADYYILE 87

RESULT 15  
AAE16325  
ID AAE16325 standard; protein; 616 AA.

```

XX AC AAE16325;
XX DT 26-MAR-2002 (first entry)
XX DE Active cellulase hybrid protein, E3/B5.
XX KW Active cellulase protein; alkalophilic; textile processing; proteinase;
XX KW detergent additive; stonewashed appearance; cotton-containing denim;
XX KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein.
XX OS Unidentified.
XX PN US6294366-B1.
XX PD 25-SEP-2001.
XX PF 19-AUG-1998; 98US-00136574.
XX PR 19-SEP-1997; 97US-00932571.
XX PA (CLRN ) CLARIANT FINANCE BVI LTD.
XX PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
XX PI Morgan H, Williams DP;
XX DR WPI; 2002-081780/11.
XX DR N-PSDB; AAD26568.
XX PT New cellulase active protein, useful in textile processing or commercial
XX PT detergents, e.g. for improving the feel or appearance of cotton-
XX PT containing fabrics, is stable under conditions of alkaline pH and
XX PT elevated temperatures.
XX PS Disclosure; Col 75-78; 61pp; English.
XX CC The present invention relates to a cellulase active protein, which is
XX CC substantially free of proteinases of native thermophilic and
XX CC alkaliphilic origin, where the cellulase active protein consists of the
XX CC CelB5 amino acid sequence. The cellulase active protein is useful for
XX CC treating cellulosic materials including cotton-containing fabrics, as
XX CC detergent additives. The cellulase active protein is also useful for
XX CC improving the feel and/or appearance of cotton-containing fabrics, for
XX CC removing surface fibers from cotton-containing knits or for imparting
XX CC stonewashed appearance to cotton-containing denims. The present proteins
XX CC are stable under condition of alkaline pH and elevated temperatures, thus
XX CC suitable for textile processing and in commercial detergents. The present
XX CC sequence is E3/B5 hybrid protein
SQ Sequence 616 AA;

Query Match 41.6%; Score 195.5; DB 5; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.3e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKXNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVTFTRDGGSTLVNCD 60
Db 1 MGSQVKVLYKXNETSASTGSIKPFVKVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 59

QY 61 WAAMGCGNIRASFGSVNPAFTADTYLQ 88
Db 60 WAQIGASNVTFNFVKLSSGVSGADYYLE 87

```

Search completed: May 11, 2004, 12:06:51  
Job time : 16.4613 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:39:45 ; Search time 14.2988 Seconds  
(without alignments)  
1738.900 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKXNDSPAGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	469	100.0	88	7	ADD22925	PA	Add22925 Acidother
2	469	100.0	89	7	ADD22924	PA	Add22924 Acidother
3	469	100.0	154	6	ABP73020	PA	ABP73020 Amino aci
4	469	100.0	762	6	ABP73022	PA	ABP73022 Amino aci
5	466	99.4	150	6	ABP73025	PA	ABP73025 Amino aci
6	466	99.4	1228	6	ABP73029	PA	ABP73029 Amino aci
7	462	98.5	88	6	ABP73018	PA	ABP73018 Amino aci
8	462	98.5	89	6	ABP73017	PA	ABP73017 Amino aci
9	462	98.5	957	6	ABP73015	PA	ABP73015 Amino aci
10	462	98.5	957	7	ADD22921	PA	Add22921 Acidother
11	436	93.0	153	6	ABP71658	PA	ABP71658 A. cellul
12	436	93.0	1121	6	ABP71656	PA	ABP71656 A. cellul
13	231.5	49.4	782	2	AAR15625	CC	Aar15625 Cellulase
14	195.5	41.7	616	2	AAV13494	CC	AAV13494 Truncated
15	195.5	41.7	616	5	AAE16325	CC	AAE16325 Active ce
16	194.5	41.5	1751	5	AAV13493	CC	AAV13493 Truncated
17	194.5	41.5	1751	5	AAE16324	CC	AAE16324 Active ce
18	191.5	40.8	1426	2	AAV13492	CC	AAV13492 Truncated
19	191.5	40.8	1426	5	AAE16323	CC	AAE16323 Active ce
20	188	40.1	477	5	AAO22444	CC	AAO22444 Protein e
21	188	40.1	496	5	AAO22443	CC	AAO22443 Protein e
22	188	40.1	677	5	AAU98063	CC	AAU98063 Bacillus
23	186	39.7	499	2	AAR42122	CC	Aar42122 NK-1 cell
24	172	36.7	700	2	AAR13227	CC	Aar13227 Novel end
25	165.5	35.3	551	2	AAW18790	CC	AAW18790 Corrected

26	165	35.2	167	2	AAR95080	PA	Aar95080 Cellulose
27	165	35.2	476	3	AAV54123	PA	Aay54123 A mamana
28	165	35.2	493	2	AAV28850	PA	Aay28850 Pectate 1
29	165	35.2	493	2	AAV43218	PA	Aay43218 Pectate 1
30	165	35.2	531	2	AAW15238	PA	Aaw15238 Scaffoldi
31	165	35.2	1853	2	AAW43108	PA	Aaw43108 C. thermo
32	155.5	33.2	1352	4	AAG63962	PA	Aag63962 Amino aci
33	150.5	32.1	1350	4	AAG63963	PA	Aag63963 Amino aci
34	130.5	27.8	1483	6	ABU23559	PA	Abu23559 Protein e
35	120.5	25.7	531	2	AAW01503	PA	Aaw01503 60 kD end
36	120.5	25.7	532	2	AAR13229	PA	Aar13229 Endogluca
37	114	24.3	162	2	AAE63634	PA	Aae63634 Cellulose
38	114	24.3	162	2	AAW90077	PA	Aaw90077 C. cellul
39	114	24.3	163	4	AAE05745	PA	Aae05745 Clostridi
40	114	24.3	256	4	AAE81128	PA	Aae81128 C17E2 Osp
41	114	24.3	256	5	AAU97869	PA	Aau97869 E. coli c
42	114	24.3	328	4	AAE05749	PA	Aae05749 Chimeric
43	114	24.3	341	4	AAE05747	PA	Aae05747 Clostridi
44	114	24.3	428	4	AAE05748	PA	Aae05748 Clostridi
45	113	24.1	190	4	AAE05746	PA	Aae05746 Clostridi

ALIGNMENTS

RESULT 1  
ADD22925  
ID ADD22925 standard; protein; 88 AA.  
XX  
AC ADD22925;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii CBD III #2.  
XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;  
KW municipal solid waste; glycoside hydrolase; avicelase.  
XX  
OS Acidothermus cellulolyticus.  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
(DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel MS;  
XX WPI; 2003-810853/76.  
XX  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX  
PS Claim 16; SEQ ID NO 5; 29pp; English.  
XX  
CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable Aviii polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase Aviii CBD III #2.



XX SQ Sequence 88 AA;  
 Query Match 100.0%; Score 469; DB 7; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-46; Indels 0; Gaps 0;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60  
 DB 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88

RESULT 2  
 ADD22924  
 ID ADD22924 standard; protein; 89 AA.  
 AC ADD22924;  
 DT 15-JAN-2004 (first entry)  
 DE Acidotherrnus cellulolyticus avicelase AvIII CBD III #1.  
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX Acidotherrnus cellulolyticus.  
 XX Key Location/Qualifiers  
 FH Misc-difference 89 /label= Unknown  
 FT  
 FT  
 XX US2003108988-A1.  
 PN 12-JUN-2003.  
 XX 18-OCT-2002; 2002US-00155400.  
 XX 28-JUL-2001; 2001US-00917376.  
 PR (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 WPI; 2003-810853/76.  
 XX New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX Claim 15; SEQ ID NO 4; 29pp; English.  
 XX The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidotherrnus  
 CC cellulolyticus avicelase AvIII CBD III #1.

XX SQ Sequence 89 AA;  
 Query Match 100.0%; Score 469; DB 7; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 4e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60  
 DB 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88

RESULT 3  
 ABP73020  
 ID ABP73020 standard; peptide; 154 AA.  
 AC ABP73020;  
 XX 03-JUN-2003 (first entry)  
 DT Amino acid sequence of the ManA carbohydrate binding domain type III.  
 DE ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
 KW food; feed; paper pulp; biofuel; mannanase.  
 XX Acidotherrnus cellulolyticus.  
 OS WO2003012110-A1.  
 PN 13-FEB-2003.  
 XX 28-JUL-2001; 2001WO-US023819.  
 PR 28-JUL-2001; 2001WO-US023819.  
 XX (MIDE ) MIDWEST RES INST.  
 PA Ding S, Adney WS, Vinzant TB, Himmel ME;  
 WPI; 2003-248182/24.  
 XX Novel thermal tolerant mannanase A polypeptide derived from Acidotherrnus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX Claim 8; Page 7; 46pp; English.  
 XX The present sequence represents a fragment of ManA, a thermostable  
 CC mannanase A polypeptide derived from Acidotherrnus cellulolyticus. ManA is  
 CC a member of the glycoside hydrolase family of enzymes. ManA is useful for  
 CC reducing hemicellulose in a starting material to simpler carbohydrate  
 CC units, and ultimately to sugars which are useful in the food, feed, paper  
 CC pulp, and biofuels industries. It is useful for the processing of food  
 CC and in food stuffs as bulking agents, and for the degradation of mannanase.  
 CC ManA is also useful to raise polyclonal and monoclonal antibodies that  
 CC are useful in purifying ManA, or detecting ManA polypeptide expression,  
 CC and as well as reagent tools for characterizing the molecular actions of  
 CC ManA polypeptides

XX SQ Sequence 154 AA;  
 Query Match 100.0%; Score 469; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-46;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60  
 DB 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSNPATPTADTYLQ 88

RESULT 4  
ABP73022  
ID ABP73022 standard; protein; 762 AA.  
XX AC  
XX ABP73022;  
XX DT 03-JUN-2003 (first entry)  
XX DE  
XX DE Amino acid sequence of the ManA polypeptide.  
XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
XX KW food; feed; paper pulp; biofuel; mannanase.  
XX XX  
XX OS Acidotherrmus cellulolyticus.  
XX PN WO2003012110-A1.  
XX PD 13-FEB-2003.  
XX PF 28-JUL-2001; 2001WO-US023819.  
XX PR 28-JUL-2001; 2001WO-US023819.  
XX PA (MIDE ) MIDWEST RES INST.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX DR WPI; 2003-248182/24.  
XX DR N-PSDB; AB277633.  
XX XX  
XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidotherrmus  
XX PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
XX PT for processing of food, and as bulking agents in food stuffs.  
XX PS Claim 3; Page 18-19; 46pp; English.  
XX XX  
XX CC The present sequence represents ManA, a thermostable mannanase A  
XX CC polypeptide derived from Acidotherrmus cellulolyticus. ManA is a member of  
XX CC the glycoside hydrolase family of enzymes. ManA is useful for reducing  
XX CC hemicellulose in a starting material to simpler carbohydrate units, and  
XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and  
XX CC biofuels industries. It is useful for the processing of food and in food  
XX CC stuffs as bulking agents, and for the degradation of mannanase. ManA is  
XX CC also useful to raise polyclonal and monoclonal antibodies that are useful  
XX CC in purifying ManA, or detecting ManA polypeptide expression, and as well  
XX CC as reagent tools for characterizing the molecular actions of ManA  
XX CC polypeptides  
XX SQ Sequence 762 AA;  
Query Match 100.0%; Score 469; DB 6; Length 762;  
Best Local Similarity 100.0%; Pred. No. 5.7e-45;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
DB 455 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 514  
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542  
RESULT 5  
ABP73025  
ID ABP73025 standard; peptide; 150 AA.  
XX AC  
XX ABP73025;  
XX DT 03-JUN-2003 (first entry)  
XX DE  
XX DE Amino acid sequence of the GuxA carbohydrate binding domain type III.

KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile.  
XX OS Acidotherrmus cellulolyticus.  
XX PN WO2003012109-A1.  
XX PD 13-FEB-2003.  
XX PF 28-JUL-2001; 2001WO-US023817.  
XX PR 28-JUL-2001; 2001WO-US023817.  
XX PA (MIDE ) MIDWEST RES INST.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
XX DR WPI; 2003-239526/23.  
XX XX  
XX PT Novel thermal tolerant GuxA polypeptide derived from Acidotherrmus  
XX PT cellulolyticus, useful for reducing cellulose in a starting material, and  
XX PT for the conversion of biomass to biofuels and biofuel additives.  
XX PS Claim 6; Page 7; 47pp; English.  
XX XX  
XX CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a  
XX CC thermostable cellulase, and is a member of the glycoside hydrolase family  
XX CC of enzymes. GuxA is useful for reducing cellulose in a starting material  
XX CC such as agricultural biomass to sugars. This is useful in biofuel  
XX CC production. GuxA is also useful in the conversion of biomass to biofuels  
XX CC and biofuel additives, in detergents, pulp and paper processing, food and  
XX CC feed processing, and in textile process. GuxA is also useful for raising  
XX CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
XX CC or detecting GuxA polypeptide expression, as well as reagent tools for  
XX CC characterizing the molecular actions of GuxA polypeptides  
XX SQ Sequence 150 AA;  
Query Match 99.4%; Score 466; DB 6; Length 150;  
Best Local Similarity 98.9%; Pred. No. 1.7e-45;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
DB 1 VSGGLKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTTRDGGSSSTLVYNC 60  
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
RESULT 6  
ABP73029  
ID ABP73029 standard; protein; 1228 AA.  
XX AC  
XX ABP73029;  
XX DT 03-JUN-2003 (first entry)  
XX DE  
XX DE Amino acid sequence of the GuxA polypeptide.  
XX KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
XX KW detergent; pulp processing; paper processing; feed processing; textile.  
XX OS Acidotherrmus cellulolyticus.  
XX PN WO2003012109-A1.  
XX PD 13-FEB-2003.  
XX PF 28-JUL-2001; 2001WO-US023817.  
XX PR 28-JUL-2001; 2001WO-US023817.  
XX PS Claim 6; Page 7; 47pp; English.

XX PA (MIDE ) MIDWEST RES INST.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
XX XX  
DR WPI; 2003-239526/23.  
DR N-PSDB; ABZ77634.  
XX XX  
PT Novel thermal tolerant GuxA polypeptide derived from Acidothermus  
PT cellulolyticus, useful for reducing cellulose in a starting biomass, and  
PT for the conversion of biomass to biofuels and biofuel additives.  
XX XX  
PS Claim 3; Page 19; 47pp; English.  
XX XX  
CC The present sequence represents a GuxA polypeptide. GuxA is thermostable  
CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
CC GuxA is useful for reducing cellulose in a starting material such as  
CC agricultural biomass to sugars. This is useful in biofuel production.  
CC GuxA is also useful in the conversion of biomass to biofuels and biofuel  
CC additives, in detergents, pulp and paper processing, food and feed  
CC processing, and in textile processes. GuxA is also useful for raising  
CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
CC or detecting GuxA polypeptide expression, as well as reagent tools for  
CC characterizing the molecular actions of GuxA polypeptides  
XX XX  
SQ Sequence 1228 AA;  
Query Match 99.4%; Score 466; DB 6; Length 1228;  
Best Local Similarity 98.9%; Pred. No. 2.3e-44;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60  
DB 584 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 643  
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671  
RESULT 7  
ABP73018  
ID ABP73018 standard; peptide; 88 AA.  
XX AC ABP73018;  
XX DT 03-JUN-2003 (first entry)  
XX DE Amino acid sequence of AvIII carbohydrate binding domain type II.  
XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX KW detergent; pulp processing; paper processing; feed processing; textile;  
XX KW cellulose.  
XX OS Acidothermus cellulolyticus.  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 89 /note= "unspecified residue"  
XX FT  
XX PN WO2003012090-A2.  
XX XX  
XX PD 13-FEB-2003.  
XX XX  
XX PF 28-JUL-2001; 2001WO-US023818.  
XX XX  
XX PR 28-JUL-2001; 2001WO-US023818.  
XX XX  
XX PA (MIDE ) MIDWEST RES INST.  
XX XX  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248177/24.  
XX XX  
PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvIII antibodies for  
PT purifying recombinant AvIII polypeptides from genetically engineered  
XX XX

PT host cells.  
XX XX  
PS Claim 6; Page 8; 44pp; English.  
XX XX  
CC The present sequence is derived from a thermostable avicelase, designated  
CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvIII  
CC antibodies that are useful in purifying recombinant AvIII polypeptides  
CC from genetically engineered host cells, in detecting AvIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvIII polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX XX  
SQ Sequence 88 AA;  
Query Match 98.5%; Score 462; DB 6; Length 88;  
Best Local Similarity 97.7%; Pred. No. 2.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60  
DB 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYRVYFTRDGGSSTLVYVNC 60  
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
RESULT 8  
ABP73017  
ID ABP73017 standard; peptide; 89 AA.  
XX AC ABP73017;  
XX DT 03-JUN-2003 (first entry)  
XX DE Amino acid sequence of AvIII carbohydrate binding domain type III.  
XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX KW detergent; pulp processing; paper processing; feed processing; textile;  
XX KW cellulose.  
XX OS Acidothermus cellulolyticus.  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 89 /note= "unspecified residue"  
XX FT  
XX PN WO2003012090-A2.  
XX XX  
XX PD 13-FEB-2003.  
XX XX  
XX PF 28-JUL-2001; 2001WO-US023818.  
XX XX  
XX PR 28-JUL-2001; 2001WO-US023818.  
XX XX  
XX PA (MIDE ) MIDWEST RES INST.  
XX XX  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248177/24.  
XX XX  
PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvIII antibodies for  
PT purifying recombinant AvIII polypeptides from genetically engineered  
XX XX  
XX PS Claim 6; Page 8; 44pp; English.  
XX XX

CC The present sequence is derived from a thermostable avicelase, designated  
 CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
 CC and is a cellulase. Aviii is useful in the conversion of biomass to  
 CC biofuels and biofuel additives. It may be useful in the production of  
 CC detergents, pulp and paper processing, food and feed processing and in  
 CC textile processes. The thermostable Aviii peptide is useful in the  
 CC degradation of cellulose, and in generating specific anti-Aviii  
 CC antibodies that are useful in purifying recombinant Aviii polypeptides  
 CC from genetically engineered host cells, in detecting Aviii polypeptide  
 CC expression, as well as a reagent tool for characterizing the molecular  
 CC actions of the polypeptide. The Aviii polynucleotide is useful as a  
 CC source of probes or primers in various diagnostic assays  
 XX  
 SQ Sequence 89 AA;

Query Match 98.5%; Score 462; DB 6; Length 89;  
 Best Local Similarity 97.7%; Pred. No. 2.5e-45;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
 DB 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
 QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAIGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
 ABP73015  
 ID ABP73015 standard; protein; 957 AA.  
 AC ABP73015;  
 XX  
 DT 03-JUN-2003 (first entry)  
 DE  
 DE Amino acid sequence of the avicelase Aviii.  
 KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose.

OS Acidothermus cellulolyticus.  
 FH Key Location/Qualifiers  
 FT Misc-difference 957 /note= "unspecified residue encoded by N"  
 FT

PN WO2003012090-A2.  
 PD 13-FEB-2003.  
 XX 28-JUL-2001; 2001WO-US023818.  
 PF 28-JUL-2001; 2001WO-US023818.  
 PR 28-JUL-2001; 2001WO-US023818.  
 XX (MIDE ) MIDWEST RES INST.  
 PA

XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 PI WPI; 2003-248177/24.  
 XX N-PSDB; ABZ77632.  
 DR

XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-Aviii antibodies for  
 PT purifying recombinant Aviii polypeptides from genetically engineered  
 PT host cells.

PS Claim 2; Page 20; 44pp; English.

XX The present sequence represents a thermostable avicelase polypeptide,  
 CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. Aviii is useful in the conversion of

CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable Aviii peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC Aviii antibodies that are useful in purifying recombinant Aviii  
 CC polypeptides from genetically engineered host cells, in detecting Aviii  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays  
 XX  
 SQ Sequence 957 AA;

Query Match 98.5%; Score 462; DB 6; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
 DB 86 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 928  
 QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
 DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 10  
 ADD22921  
 ID ADD22921 standard; protein; 957 AA.  
 XX  
 AC ADD22921;  
 XX  
 DT 15-JAN-2004 (first entry)  
 DE  
 DE Acidothermus cellulolyticus avicelase Aviii.  
 XX enzyme; Aviii; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.

OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 957 /label= Unknown  
 FT /note= "Encoded by N"  
 FT

PN US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 PI WPI; 2003-810853/76.  
 XX N-PSDB; ADD22922.  
 DR

XX New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 1; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable Aviii polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII.

XX  
 SQ Sequence 957 AA;

Query Match 98.5%; Score 462; DB 7; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.8e-44;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVYNC 60  
 DB 869 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVYNC 928  
 QY 61 WAAICGNGIRASFGSVNPATPTADTYLQ 88  
 DB 929 WAAICGNGIRASFGSVNPATPTADTYLQ 956

RESULT 11

ABP71658  
 ID ABP71658 standard; protein; 153 AA.

AC ABP71658;

DT 29-MAY-2003 (first entry)

XX A. cellulolyticus Gux1 protein CBD\_III domain fragment.

XX Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.

XX Acidothermus cellulolyticus.

XX WO2003012095-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023820.

XX 28-JUL-2001; 2001WO-US023820.

XX (MIDE ) MIDWEST RES INST.

XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;

XX WPI; 2003-300494/29.

XX New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.

XX Claim 2; Page 7; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase CBD\_III  
 CC domain fragment

XX Sequence 153 AA;

Query Match 93.0%; Score 436; DB 6; Length 153;  
 Best Local Similarity 96.4%; Pred. No. 4.9e-42;  
 Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 VKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVYNC 64  
 DB 4 LKAQYKNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVYNC 63

QY 65 GCGNIRASFGSVNPATPTADTYLQ 88

DB 64 GCGNIRASFGSVNPATPTADTYLQ 87

RESULT 12

ABP71656

ID ABP71656 standard; protein; 1121 AA.

XX ABP71656;

XX 29-MAY-2003 (first entry)

XX A. cellulolyticus Gux1 protein.

XX Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; enzyme.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT Peptide 1..34

FT Protein /note= "potential signal peptide"

FT Misc-difference 228 /note= "mature protein"

FT /note= "encoded by CG"

XX WO2003012095-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023820.

XX 28-JUL-2001; 2001WO-US023820.

XX (MIDE ) MIDWEST RES INST.

XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;

XX WPI; 2003-300494/29.

XX N-PSDB; ABZ76162.

XX New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.

XX Claim 3; Page 18-19; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase

XX Sequence 1121 AA;

AD	AAV13494 standard; protein; 616 AA.
ID	AAV13494
XX	XX
XX	AC
XX	AAV13494;
XX	XX
XX	30-JUL-1999 (first entry)
DT	XX
XX	Truncated cellulase Cel E3/B5.
DE	XX
XX	XX
KW	Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW	Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW	cotton-containing fabric; stonewashing.

XX	unidentified.
XX	
PN	EP921188-A2.
XX	
XX	
PD	09-JUN-1999.
XX	
XX	
PF	15-SEP-1998; 98EP-00810919.
XX	
XX	
PR	19-SEP-1997; 97US-00932571.
XX	
XX	(CLRN ) CLARIANT FINANCE BVI LTD.
PA	
PI	Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI	Morgan H, Williams DP;
XX	
DR	WPI: 1999-315403/27.
DR	N-PSDB; AAX55660.
XX	
XX	New truncated cellulase proteins, useful in detergents and for producing
PT	'stonewashed' denim.
PT	
XX	
PS	Claim 7; Page 42-43; 65pp; English.
XX	
CC	The invention relates to a recombinant cellulase active protein free of
CC	proteinases of native thermophilic and alkalophilic origin, comprising
CC	the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC	Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC	length sequences, or functional equivalents. Cel B5 extend from amino

acid A4911 to N1424 of N1426, and Cei B4/5 extends from amino acid G635 to N1426 in the sequence shown in AA13492; Cei E1 extends from amino acid Y39 to D481, Cei E1/2 extends from Y39 to G635, Cei E1/2/3 extends from Y39 to G812, Cei B6 extends from amino acid V1233 to K1751 and the stability region extends from amino acid S482 to G635 in the sequence shown in AA13493; Cei E3/B5 is shown in AA13494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic materials including cotton-containing fabrics. They are especially useful for preventing redeposition of colorant during stonewashing, and for processing of textiles where cellulose breakdown is required. The new truncated enzymes show reduced redeposition of dye compared to using non-truncated cellulase compositions

```

Query Match      41.7%; Score 195.5; DB 2; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.3e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps
Qy      1 VSGGVKQVKNNDSPGDNQIKPGQLVNTGSSVDLSTVTVRVYFTPRDGGSSSTLVNCD 60
      : ||| ||| :: : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1 MGSGVKLVLYKNNETSASTGSRPWFKIVNGGSSVDLSRVRKIRVYWTVDGDKPQSAV-CD 59
Qy      61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      60 WAQIGASNVTFNFKVLLSGVSGADYYLE 87
RESULT 15
AAE16325
ID AAE16325 standard; protein: 616 AA

```

XX AAEL6325;  
 XX AC  
 XX DT  
 XX DE  
 XX DE Active cellulase hybrid protein, E3/B5.  
 XX KW Active cellulase protein; alkalophilic; textile processing; proteinase;  
 KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein.  
 XX OS Unidentified.  
 XX PN USG294366-B1.  
 XX PD 25-SEP-2001.  
 XX PF 19-AUG-1998; 98US-00136574.  
 XX PR 19-SEP-1997; 97US-00932571.  
 XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX DR WPI; 2002-081780/11.  
 XX DR N-PSDB; AAD26568.  
 XX PT New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX PS Disclosure; Col 75-78; 61pp; English.  
 XX CC The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the  
 CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is E3/B5 hybrid protein  
 XX SQ Sequence 616 AA;  
 Query Match 41.7%; Score 195.5; DB 5; Length 616;  
 Best Local Similarity 43.2%; Pred. No. 1.3e-13;  
 Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGQLVNTGSSVDLSVTIVRYWFTDRGSSSTLVYNC 60  
 Db : ||| : ||| :  
 1 MCGGVKVLKNETASTGSIPTPFKIVNGSSVDLSRVKIRYWTVDGKPSAV-CD 59  
 QY 61 WAAMCGNIRASFGSVNPATPTATY 88  
 Db :  
 60 WAQIGASNVTFNFVKLSSGVSGADYYLE 87

Search completed: May 11, 2004, 12:06:51  
 Job time : 14.2988 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:51:45 ; Search time 4.1265 Seconds  
(without alignments)  
2051.340 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKVOYKNDSPAGNQ.....TRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	191	40.7	1711	2 T31337	1,4-beta-glucanase
2	190.5	40.6	1779	2 T31085	xylanase - Caldice
3	190	40.5	508	2 G63593	cellulase (EC 3.2.
4	190	40.5	508	2 A26874	cellulase (EC 3.2.
5	188	40.1	499	2 JN0111	cellulase (EC 3.2.
6	186.5	39.8	1331	2 A48954	mannan endo-1,4-be
7	186.5	39.8	1742	2 T17120	cellulase (EC 3.2.
8	186	39.7	499	2 A27198	cellulase (EC 3.2.
9	182.5	38.9	915	2 A43802	cellulase (EC 3.2.
10	182.5	38.9	1039	2 S02711	cellulase (EC 3.2.
11	180.5	38.5	145	2 A41897	cellulase homolog
12	180	38.4	486	2 I40548	bifunctional cellu
13	172	36.7	700	2 B41897	cellulase (EC 3.2.
14	165.5	35.3	879	2 A47704	endoglucanase I (E
15	165	35.2	1854	2 S36859	cipA protein - Clo
16	149.5	31.9	505	2 S39962	endoglucanase - Er
17	147.5	31.4	504	2 S54744	cellulase (EC 3.2.
18	141.5	30.2	986	2 S12021	thermoactive cellu
19	130.5	27.8	586	2 PC5006	scaffolding proteoi
20	126.5	27.0	1483	2 C97012	probably celluloso
21	126.5	27.0	1162	2 T30433	scaffolding protei
22	117	24.9	1230	2 T30433	cellulose 1,4-beta
23	114	24.3	1848	2 A44140	cellulose-binding
24	75.5	16.1	618	2 T08685	hypothetical prote
25	73.5	15.7	547	2 T25478	hypothetical prote
26	73.5	15.7	1428	2 AC2224	hypothetical prote
27	70	14.9	5188	2 B85547	probable RTX famil
28	70	14.9	5291	2 F90696	hypothetical prote
29	68.5	14.6	574	2 A69196	cell surface glyco

30	67	14.3	271	2 H72684	hypothetical prote
31	66.5	14.2	247	2 JC7201	thauatin-like pro
32	66	14.1	535	2 T47790	hypothetical prote
33	65.5	14.0	751	2 T40462	ser-lys rich hypot
34	65.5	14.0	1435	2 T01075	polyprotein - hepa
35	65	13.9	523	2 C70717	probable purH prot
36	64	13.6	1222	2 G72614	probable reverse g
37	64	13.6	1571	2 S50669	hypothetical prote
38	64	13.6	1594	2 T30549	hemolysin - rabbit
39	63.5	13.5	398	2 H70393	hemolysin - Aquife
40	63.5	13.5	2403	2 A59386	sanko - human
41	62.5	13.3	269	2 H95976	endo-beta-1,3-1,4-
42	62.5	13.3	454	2 AB2512	replicative DNA he
43	62.5	13.3	802	1 B44390	protein-tyrosine-p
44	62.5	13.3	4936	2 AH2515	hypothetical prote
45	62	13.2	122	2 E37267	Ig heavy chain V r

#### ALIGNMENTS

##### RESULT 1

T31337

1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C;Species: Anaerocellum thermophilum

C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C;Accession: T31337

R;Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144: 457-465, 1998  
A;Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from omain.

A;Reference number: Z21003; MUID:98154434; PMID:9493383

A;Accession: T31337

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1711 <ZVS>

A;Cross-references: EMBL:Z86105; NID:el071329; PID:e350354; PIDN:CAB06786.1

C;Genetics:

A;Gene: cela

C;Keywords: glycosidase; hydrolase

Query Match

Best Local Similarity 40.7%; Score 191; DB 2; Length 1711;

Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-VKVOYKNDSPAGNQIKPGLQVNTGSSVDLTSTVRYWETRDGSGSTLYVNC 59

DB 683 VAGGQIKVLYANKETNSTTIRPWLVKVTGSSIDLSRVTIRYWTVDGKAQSAIS- 741

QY 60 DWAAWCGNIRASFSGSVNPATPTADTYLQ 88

DB 742 DWAIQASNVTFKVLSSVSGADYILE 770

##### RESULT 2

T31085

xylanase - Caldicellulosiruptor sp.

C;Species: Caldicellulosiruptor sp.

C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C;Accession: T31085

R;Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL Data Library, December 1997

A;Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.

A;Reference number: Z20972

A;Accession: T31085

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1779 <MOR>

A;Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1

C;Genetics:

A;Note: xync

Query Match

Score 190.5; DB 2; Length 1779;





```

|||||: |::: |||||:
Db 419 QMGCGNLTHKFTVTLHKPKQGADTYLE 444

RESULT 5
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
C:Accession: JN0111
R:Park, S.H.; Kim, H.K.; Pack, M.Y.
Agric. Biol. Chem. 55, 441-448, 1991
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE61
A:Reference number: JN0111; MUID:91299280; PMID:1368694
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:D01057; NID:G216387; PIDN:BA00859.1; PID:d1001323; PID:G216388
A:Note: The authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 40.1%; Score 188; DB 2; Length 499;
Best Local Similarity 40.0%; Pred. No. 6.7e-12;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDRDGGSSTLVNCWAA 63
DB 353 GISVQYRAGDSMSNIRPOLQIKNGNTTVLDKDVARTWYNAKNGQNV--DCDYAQ 410

QY 64 MCGGNIRASFGSNPATPTADTYLQ 88
DB 411 LCGGNVTKFVTLHKPKQGADTYLE 435

RESULT 6
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139; PMID:1476429
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GB>
A:Cross-references: GB:L01257; NID:G144290; PIDN:AAA71887.1; PID:G144291
A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBI:P121577)
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQHRRQ' <LUE>
A:Cross-references: EMBL:M36063; NID:G144292; PIDN:AAA72861.1; PID:G144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.8%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

```

```

QY 2 SGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDRDGGSSTLVNCW 61
DB 364 SQGIKVLVYANKETNSTNTIRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 422

QY 62 AAMGCCNIRASFGSNPATPTADTYLQ 88
DB 423 AQIGASNVTFKFKLSSSVSGADTYLE 449

```

## RESULT 7

```

T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophi
A:Reference number: Z18698; MUID:95336703; PMID:7612247
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:G537499; PID:G537500; PIDN:AAA91086.1
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:G144292; PIDN:AAA72860.1; PID:G144293
C:Genetics:
A:Gene: cella
C:Keywords: glycosidase; hydrolase

```

```

Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

```

```

QY 2 SGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTDRDGGSSTLVNCW 61
DB 704 SQGIKVLVYANKETNSTNTIRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 762

QY 62 AAMGCCNIRASFGSNPATPTADTYLQ 88
DB 763 AQIGASNVTFKFKLSSSVSGADTYLE 789

```

## RESULT 8

```

A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397; PMID:3106035
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAB>
A:Cross-references: GB:M28332; NID:G142670; PIDN:AAA22307.1; PID:G142671
A:Experimental source: strain IFO3034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

```

```
F;29-1039/Product: cellulase #status predicted <MAT>
F;72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match          38.9%; Score 182.5; DB 2; Length 1039;
Best Local Similarity 41.9%; Pred. No. 5.5e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GGVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTRDGGSTLVYNCDWA 62
Db   ||| :|| ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
420 GQIKVLANKETNSTTIRPWLVKVVNSGSSIDLSRVITRIVYTVDGERAQSAVS-DWA 478

QY 63 AMCGNIRASFGSVNPATPTADTYLQ 88
Db   :|| :|| :::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
479 QIGASNVTTFKVKLSLSSVGADYYLE 504

RESULT 11
A41897
cellulase homolog - Bacillus lautus (fragment)
C;Species: Bacillus lautus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C;Accession: A41897; S27498
R;Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A;Title: cels from Bacillus lautus FL236 encodes a novel cellulose-binding endo
A;Reference number: A41897; UID:92276330; PMID:1592807
A;Accession: A41897
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-145 <HAN>
A;Cross-references: EMBL:M76598; NID:g142661; PIDN:AAA22302.1; PID:g142662
A;Experimental source: PL236
A;Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match          38.5%; Score 180.5; DB 2; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.1e-11;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

QY 7 VOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTRDGGSTLVYNCDWAAMGC 66
Db   .||:|. :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1 LQYRAADTAADNQIKPFNKNGTSVADLSLTIKTRYFTKDGSAAVNGW-IDWAQLGG 59

QY 67 GNIRASFGSVNPATPTADTYLQ 88
Db   ||::||| |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
60 SNIQISFG--NHTGTNSDTYE 79

RESULT 12
I40548
bifunctional cellulase precursor - Bacillus sp.
C;Species: Bacillus sp.
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C;Accession: I40548
R;Han, S.J.; Yoo, Y.-J.; Kang, H.S.
J. Biol. Chem. 270, 26012-26019, 1995
A;Title: Characterization of a bifunctional cellulase and its structural gene:
A;Reference number: I40548; UID:96029707; PMID:7592793
A;Accession: I40548
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-496 <RES>
A;Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576
C;Genetics:
A;Gene: cel

Query Match          38.4%; Score 180; DB 2; Length 486;
Best Local Similarity 38.8%; Pred. No. 4.5e-11;
Matches 33; Conservative 18; Mismatches 32; Indels 2; Gaps 1;

QY 4 GUVKVOYKNDSAPGDNQIKPGLQLVNTGSSVDLSVTIVRYWFTRDGGSTLVYNCDWAA 63
Db   ||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353 GISVQYRAGDGKNSNQIRPQIKNGNGTTVDLKDVARTARYWNNAKKGNL--DCDYEQ 410
```

```
QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
DB 411 LGCGNVSHVTTLHKPKQGAADTYLE 435

RESULT 13
B41897
cellulase (EC 3.2.1.4) - Bacillus lautus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus lautus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999
C:Accession: B41897; S27499
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: ccla from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: B41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663
A:Experimental source: PL236
A>Note: sequence extracted from NCBI backbone (NCBIP:104605)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 4.5e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTFDGSSSTLVYNC 59
DB 549 VNSDLVVOYKDGDRNATDNOIKPHFNQNGTSVPDLSSUTLRYFTKD-SSAAMNGWI 607

QY 60 DWAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 608 DWAKLGGNSIQISFGNHGA--DSDTYAE 634

RESULT 14
A47704
endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A47704
R:Hazlewood, G.P.; Davidon, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.
J. Gen. Microbiol. 139, 307-316, 1993
A:Title: Gene sequence and properties of CellI, a family E endoglucanase from Clostridium
A:Reference number: A47704; MUID:93171873; PMID:8436949
A:Accession: A47704
A>Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-879 <HAZ>
A:Cross-references: GB:L04735; NID:g144807; PIDN:AA20892.1; PID:g144808
A>Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)
C:Keywords: glycosidase; hydrolase

Query Match 35.3%; Score 165.5; DB 2; Length 879;
Best Local Similarity 33.0%; Pred. No. 2.7e-09;
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTFDGSSSTLVYNC 60
DB 736 IKGEVVLQYANGAGATSNINPRKIINNGTKAINLSDVKIRYYTKEGGSQNFW-CD 794

QY 61 WAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 795 WSSAGNSNVGTGNFFNLSSPKEGADTCL 822

RESULT 15
```

```
S36859
cipA protein - Clostridium thermocellum
N:Alternate names: probable cellulosome protein large chain SL
C:Species: Clostridium thermocellum
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S36859; S33527; S25767; S28659; T18261
R:Gerengross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: EMBL:L08665
R:Gerengross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615,1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1768, 'R', 1770-1854
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18261
A>Status: preliminary; translated from GS/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipA

Query Match 35.2%; Score 165; DB 2; Length 1854;
Best Local Similarity 36.1%; Pred. No. 6.9e-09;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKND-SAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTFDGSSSTLVYNC 60
DB 365 VSGNLKVFYNPSDPTTNSINPQFKVTNTGSSAIDLKSLTLRYYYTVDGQKQOTFW-CD 423

QY 61 WAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 424 HAAIGTSGNSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: May 11, 2004, 12:10:26
Job time : 4.1265 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:41:35 ; Search time 2.39913 Seconds  
(without alignments)  
1909.933 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	40.5	499	1 GUN1_BACSU	P07983 bacillus su
2	190	40.5	499	1 GUN2_BACSU	P10475 bacillus su
3	188	40.1	499	1 GUN3_BACSU	P23549 bacillus su
4	186.5	39.8	1331	1 MANB_CALSA	P22533 caldocellum
5	186.5	39.8	1742	1 GUN4_CALSA	P22534 caldocellum
6	182.5	38.9	1039	1 GUNB_CALSA	P10474 c endogluc
7	180.5	38.5	145	1 YCEA_PAEIA	P29718 paenibacill
8	172	36.7	700	1 GUN4_PAEIA	P02934 clostridium
9	165.5	35.3	879	1 GUN1_CLOTM	Q02933 clostridium
10	165	35.2	772	1 CIPB_CLOTM	Q01866 clostridium
11	165	35.2	1853	1 CIPB_CLOTM	Q06851 clostridium
12	149.5	31.9	444	1 GUNV_ERWCA	Q59394 erwinia car
13	149.5	31.9	505	1 GUNV_ERWCA	Q47096 erwinia car
14	147.5	31.4	504	1 GUNV_ERWCA	Q59395 erwinia car
15	147.5	31.4	914	1 GUX2_CLOS	P50900 clostridium
16	141.5	30.2	985	1 GUNZ_CLOS	P23659 clostridium
17	114	24.3	1848	1 CBPA_CLOCL	P38058 clostridium
18	67.5	14.4	118	1 HV39_MOUSE	P01809 mus musculu
19	67.5	14.4	215	1 FLA1_METVA	P95316 methanococc
20	66.5	14.2	246	1 TP1A_MALDO	Q9fs97 malus domes
21	65	13.9	523	1 PUR9_MYCTU	P17553 h bifunctio
22	64.5	13.8	556	1 ILVD_THETN	Q8rdj9 thermoanear
23	64	13.6	1571	1 ATCS_YEAST	P32660 saccharomyc
24	63.5	13.5	247	1 FLA1_THEVO	P57719 thermoplasm
25	63.5	13.5	465	1 TY3H_SCHMA	O17446 schistosoma
26	63	13.4	953	1 O106_HUMAN	Q9upv9 homo sapien
27	63	13.4	1581	1 PRB_HUMAN	Q15648 h peroxisom
28	62.5	13.3	269	1 EXOK_RHIME	P33693 rhizobium m
29	62.5	13.3	802	1 PTN8_MOUSE	P29352 mus musculu
30	62	13.2	950	1 STL2_MOUSE	P09815 mus musculu
31	62	13.2	1210	1 ICEN_PSEPL	P09815 pseudomonas
32	61.5	13.1	252	1 HIS6_RHOSH	P50937 rhodobacter
33	61.5	13.1	316	1 MUCB_PSEAE	P38108 pseudomonas

34	61.5	13.1	328	1 HAIQ_MOUSE	P14428 mus musculu
35	61.5	13.1	368	1 HAIW_MOUSE	P03991 mus musculu
36	61.5	13.1	489	1 GATA_CAUCR	Q9a510 caulobacter
37	61.5	13.1	678	1 YIHQ_ECOLI	P32138 escherichia
38	61	13.0	120	1 HV50_MOUSE	P06329 mus musculu
39	61	13.0	338	1 RTCA_ECOL6	O8fca8 escherichia
40	61	13.0	338	1 RTCA_ECOLI	P46849 escherichia
41	61	13.0	342	1 RTCA_ECO57	P58127 escherichia
42	60.5	12.9	1157	1 C9CA_BACTO	Q45733 bacillus th
43	60	12.8	312	1 LECF_ALEAU	P18891 aleuria aur
44	60	12.8	1226	1 PAT2_CAEEL	P34446 caenorhabdi
45	59.5	12.7	261	1 COTN_BACSU	P54507 bacillus su

#### ALIGNMENTS

RESULT 1  
GUN1\_BACSU  
ID GUN1\_BACSU STANDARD; PRT; 499 AA.  
AC P07983;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase).  
GN BGLC OR GLD.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DLG;  
RX MEDLINE=87194581; PubMed=3106328;  
RA Robson L.M., Chambliss G.H.;  
RL "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";  
RL J. Bacteriol. 169:2017-2025(1987).  
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -|- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; M16185; AAA22496.1; ALT\_INIT.

PIR; A26874; A26874.

HSSP; O85465; 1A3H.

InterPro; IPR001956; CBD 3.

InterPro; IPR008965; Cellul\_bind.

InterPro; IPR001547; Glyco\_hydro\_5.

Pfam; PF00942; CBM\_3; 1.

Pfam; PF00150; cellulase; 1.

ProDom; PD001947; CBD 3; 1.

PROSITE; P500559; GLYCOSYL\_HYDROL\_F5; 1.

Cellulose degradation; Hydrolase; Glycosidase; Signal.

SIGNAL 1 29

CHAIN 30 499 ENDOGLUCANASE.

FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).

FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).

FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).

SQ SEQUENCE 499 AA; 55187 MW; 339D04BE95A63BE1 CRC64;

Query Match 40.5%; Score 190; DB 1; Length 499;

Best Local Similarity 43.0%; Pred. No. 1.3e-13;

Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

OY 4 GVKVQKNNDSAPGDNQIKPGIQLVNTGSSVDLSTVTYVWF-TRDGGSTLTVNCDA 62

```

Db 353 GVSQYKAGDGVNSQIRPOLHKNNGNATVDLKDVTARYWYVKNKNGQN---FDCDYA 409
QY 63 AMGCGNIRASFGSVNPNPATPTADTYLQ 88
Db 410 QMCGCNLTHKFTVTLHKPKQGADTYLE 435

RESULT 2
GUN2_BACSU
ID GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCASE) (Cellulase).
GN BGIC OR GLD OR EGLS OR ESU18130.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA115;
RX MEDLINE=87066783; PubMed=3024130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moranelli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degrees region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter.";
RL Microbiology 142:3097-3101(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartori M.G., Bessieres P., Bolotin A., Borchert S.,
RA Bourissier R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dunsthorst A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

```

```

RA Takeuchi M., Takakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; CAA82317.1; -
DR EMBL; X04689; CAA28392.1; -
DR EMBL; X67044; CAA47429.1; -
DR EMBL; Z73234; CAA97610.1; ALT_INIT.
DR EMBL; Z99113; CAB13696.1; ALT_INIT.
DR FIR; G69593; G69593.
DR HSP; O85465; 1A3H.
DR Subtilist; BG10437; bglC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAE2 CRC64;

Query Match 40.5%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred. No. 13e-13;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDASPGDNOIRPGLQVNTGSSVDLSTVTVRYWF-TRDGGSTLYVNCDA 62
Db 353 GISVQYRAGDGSNNQIRPOLQIKNGNTVDLKDVTARYWYVKNKNGQN---FDCDYA 409
QY 63 AMGCGNIRASFGSVNPNPATPTADTYLQ 88
Db 410 QMCGCNLTHKFTVTLHKPKQGADTYLE 435

RESULT 3
GUN3_BACSU
ID GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)

```

DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).  
GN BGLC.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BS616;  
RX MEDLINE=91299280; PubMed=1368694;  
RA Park S.H., Kim H.K., Pack M.Y.;  
RT "Characterization and structure of the cellulase gene of Bacillus subtilis BSE616";  
RL Agric. Biol. Chem. 55:441-448(1991).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; D01057; BAA0859.1; -;  
DR PIR; JN0111; JN0111.  
DR HSP; O85465; I43H.  
DR InterPro; IPR001956; CBD 3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001547; Glyco\_Hydro\_5.  
DR Pfam; PF00942; CBM\_3; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR ProDom; PD001947; CBD 3; 1.  
DR ProSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 499 ENDOGLUCANASE.  
FT ACT SITE 169 169 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).  
SQ SEQUENCE 499 AA; 55169 MW; 2E21E3DBBACA04 CRC64;  
Query Match 40.1%; Score 188; DB 1; Length 499;  
Best Local Similarity 40.0%; Pred. No. 2.2e-13;  
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
QY 4 GYKVOYKNDAPGDNQIKPGLQVLVNTGSSVDLSLSTVTVYVFTDRDGGSSTLVYVNCWNA 63  
Db 353 GSVQYRAGDSGMSNRIPQLQIKNGNTVTLKDVTVYVYNAKRGQNV--DCDYAQ 410  
QY 64 MCGNIRASFGSVNPATPTADTYLQ 88  
Db 411 LCGGNVTKFVTLHKPKQGADTYLE 435  
RESULT 4  
MANE\_CALSA  
ID MANE\_CALSA STANDARD; PRT; 1331 AA.  
AC P22533;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].  
GN MANA.  
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;

OC Caldicellulosiruptor.  
OX NCBI\_TaxID=44001;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93119139; PubMed=1476429;  
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;  
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme";  
RL Appl. Environ. Microbiol. 58:3864-3867(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91247819; PubMed=2039230;  
RA Luehli E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'";  
RL Appl. Environ. Microbiol. 57:694-700(1991).  
CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees Celsius.  
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; L01257; AAA71887.1; -;  
DR EMBL; M36063; AAA72861.1; -;  
DR PIR; A48954; A48954.  
DR HSP; O06851; INBC.  
DR InterPro; IPR001956; CBD 3.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001547; Glyco\_Hydro\_5.  
DR Pfam; PF00942; CBM 3; 2.  
DR Pfam; PF00150; cellulase; 1.  
DR ProDom; PD001947; CBD 3; 2.  
DR ProSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
KW Hydrolase; Glycosidase; Cellulose degradation; Signal;  
FT SIGNAL 1 41  
FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.  
FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).  
FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).  
FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).  
FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).  
FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).  
FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).  
FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).  
FT ACT SITE 162 162 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
FT CONFLICT 338 338 T -> P (IN REF. 2).  
FT CONFLICT 340 346 TPTPT -> RQHQHQ (IN REF. 2).  
SQ SEQUENCE 1331 AA; 146892 MW; PFC5A1BB8DF0E0 CRC64;  
Query Match 39.8%; Score 186.5; DB 1; Length 1331;  
Best Local Similarity 42.5%; Pred. No. 1e-12;  
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;  
QY 2 SGVKVQYKNDAPGDNQIKPGLQVLVNTGSSVDLSLSTVTVYVFTDRDGGSSTLVYVNCW 61



```

Db 364 SQGIKVLANKETNTTIRPWLKVNSSGSSIDLSTRTIRYWTVDGERAQSAIS-DW 422
QY 62 AAMGCGNIRASFGSVNPAFTADTYLQ 88
Db 423 AQIGASNVTFKVLSSSVSGADYILE 449

RESULT 5
GUNA_CALSA
ID GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=203230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: The linker region (also termed "hinge") may be a potential
CC site for proteolysis.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L32742; AAA91086.1; -
CC EMBL; M36063; AAA72860.1; -
CC EMBL; L01257; -; NOT_ANNOTATED_CDS.
CC PIR; T17120; T17120.
CC HSP; P26221; 1TP4.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR000556; Glyco_hydro_48.
CC InterPro; IPR001701; Glyco_hydro_9.
CC InterPro; IPR008928; Glyco_trans_6hp.
CC Pfam; PF00942; CEM_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC PRINTS; PR00844; GLHYDLASE48.
CC ProDom; PD001947; CBD_3; 2.

```

```

DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 39.8%; Score 186.5; DB 1; Length 1742;
Best Local Similarity 42.5%; Pred. No. 1.4e-12;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVFTDGGSTLVYNCNW 61
Db 704 SQGIKVLANKETNTTIRPWLKVNSSGSSIDLSTRTIRYWTVDGERAQSAIS-DW 762
QY 62 AAMGCGNIRASFGSVNPAFTADTYLQ 88
Db 763 AQIGASNVTFKVLSSSVSGADYILE 789

RESULT 6
GUNB_CALSA
ID GUNB_CALSA STANDARD; PRT; 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellobiohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exoglucanase and endocellulase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
CC F (family 10 of glycosyl hydrolases).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC modified and this statement is not removed. Usage by and for commercial

```



```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13602; CAA31936.1; -.
DR PIR; S02711; S02711.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Repeat;
KW Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039
FT DOMAIN 376 416
FT DOMAIN 417 570
FT DOMAIN 571 618
FT ACT_SITE 177 177
FT ACT_SITE 285 285
FT ACT_SITE 792 792
FT ACT_SITE 1039 1039
FT SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;
Query Match 38.9%; Score 182.5; DB 1; Length 1039;
Best Local Similarity 41.9%; Pred. No. 2.1e-12;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
QY 3 GGKVKVQNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCDWA 62
DB 420 QIKVLVANKETTTTIRPWLKVNSSGSSIDLSRTVIRYVYTVDGERAQSAVS-DWA 478
QY 63 AMGCNIRASFGSVNPATPTADTYLQ 88
DB 479 QIGASNTVTFKVLSSVSGADYYLE 504
RESULT 7
YCEA_PAELA STANDARD; PRT; 145 AA.
AC P29718;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in cella 5' region (Fragment).
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Joergensen P.L., Diderichsen B.;
RT "cella from Bacillus lautus PL236 encodes a novel cellulose-binding
end-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76588; AAA22302.1; -.
DR PIR; A41897; A41897.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700
FT ACT_SITE 213 213
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EEB0F CRC64;
Query Match 36.7%; Score 172; DB 1; Length 700;
PIR; A41897; A41897.
```

```
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER 1
FT SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;
Query Match 38.5%; Score 180.5; DB 1; Length 145;
Best Local Similarity 43.9%; Pred. No. 3.4e-13;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;
QY 7 VOYKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCDWAAMGC 66
DB 1 LQYRAADTNAADNQIKPSFNKNGTSAVDLSTLKIRYFTKGSAAVNGW-IDWAQLGG 59
QY 67 GNIRASFGSVNPATPTADTYLQ 88
DB 60 SNIQISFG--NHTGTSNTDYE 79
RESULT 8
GUNA_PAELA STANDARD; PRT; 700 AA.
ID GUNA_PAELA
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase).
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "cella from Bacillus lautus PL236 encodes a novel cellulose-binding
end-beta-1,4-glucanase."
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76588; AAA22303.1; -.
DR PIR; B41897; B41897.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700
FT ACT_SITE 213 213
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EEB0F CRC64;
Query Match 36.7%; Score 172; DB 1; Length 700;
PIR; B41897; B41897.
```

795 WSASGNSRVICAFNDSSPFRGRDUCLE 822

RESULT 9  
GUNI\_CLOTHM STANDARD; PRT; 879 AA.

ID ID GUNI\_CLOTHM STANDARD; PRT; 879 AA.  
AC Q02934;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)  
DE (Cellulase I).  
GN CELI.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.  
RC STRAIN=NCIB 10682;  
RX MEDLINE=93171873; PubMed=8436949;  
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,  
Gilbert H.J.;  
RT "Gene sequence and properties of Celli, a family E endoglucanase from  
Clostridium thermocellum.";  
RL J. Gen. Microbiol. 139:307-316(1993).  
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOLYOLYSIS OF 1,4-BETA-  
GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PATHWAY: Cellulose degradation.  
CC -!- SIMILARITY: Belongs to cellulase family E (family 9 of glycosyl  
hydrolases).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; L04735; AAA20892.1; --  
PIR; A47704; A47704.  
HSP; P26221; ITF4.  
InterPro; IPR001956; CBD\_3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR001701; Glyco\_Hydro\_9.  
InterPro; IPR008928; Glyco\_trans\_6hp.  
Pfam; PF00942; CBM\_3; 2.  
Pfam; PF00759; Glyco\_hydro\_9; 1.  
ProDom; PD001947; CBD\_3; 1.  
ProSITE; PS00592; GLYCOSYL\_HYDROL\_F9\_1; 1.  
ProSITE; PS00698; GLYCOSYL\_HYDROL\_F9\_2; 1.  
Cellulose degradation; Hydrolase; Glycosidase; Signal.  
KW SIGNAL  
FT CHAIN 1 55  
FT DOMAIN 56 879 ENDOGLUCANASE I.  
FT CATALYTIC.  
FT DOMAIN 56 518  
FT DOMAIN 729 879 CELLULOSE-BINDING (BY SIMILARITY).  
FT ACT\_SITE 448 448 BY SIMILARITY.  
FT ACT\_SITE 486 486 BY SIMILARITY.  
FT ACT\_SITE 495 495 BY SIMILARITY.  
FT SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;  
SQ

RESULT 10  
CIPB\_CLOTHM STANDARD; PRT; 772 AA.

ID ID CIPB\_CLOTHM STANDARD; PRT; 772 AA.  
AC Q01866;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein S1/SL)  
DE (Cellulose integrating protein B) (Fragment).  
GN CIPB.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]\_TaxID=1515;  
RP SEQUENCE FROM N.A.  
RC STRAIN=YS;  
RX MEDLINE=93146373; PubMed=1490597;  
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,  
Gilbert H.J.;  
RT "Identification of the cellulose-binding domain of the cellulosome  
subunit S1 from Clostridium thermocellum YS.";  
RL FEMS Microbiol. Lett. 78:181-186(1992).  
CC -!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT  
PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE  
REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS  
PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.  
CC -!- SUBCELLULAR LOCATION: Cell surface.  
CC -!- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY  
THE CATALYTIC COMPONENTS OF THE CELLULOSE.  
CC -!- SIMILARITY: Contains at least 3 cohesin domains.  
CC -!- SIMILARITY: Contains 2 dockerin domains.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; X68233; CAA48312.1; --  
HSP; Q06851; INBC.  
InterPro; IPR001956; CBD\_3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR002102; Cohesin.  
InterPro; IPR002105; Dockerin\_1.  
InterPro; IPR002048; EF-hand.  
Pfam; PF00942; CBM\_3; 1.  
Pfam; PF00963; Cohesin; 3.  
Pfam; PF04004; Dockerin\_1; 2.  
ProDom; PD001947; CBD\_3; 1.  
ProSITE; PS00018; EF\_HAND; UNKNOWN 1.  
ProSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
Cellulose degradation; Cell wall; Glycoprotein; Repeat.  
NON\_TER 1 1  
DOMAIN <1 80 COHESIN 1.  
FT LINKER (PRO/THR-RICH).  
FT 81 93  
FT DOMAIN 93

```

FT DOMAIN 94 240 COHESIN 2.
FT DOMAIN 241 272 LINKER (PRO/THR-RICH).
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 LINKER (PRO/THR-RICH).
FT DOMAIN 462 607 COHESIN 3.
FT DOMAIN 710 733 DOCKERIN 1.
FT DOMAIN 743 766 DOCKERIN 2.
SQ SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 35.2%; Score 165; DB 1; Length 772;
Best Local Similarity 36.1%; Pred. No. 1.4e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDAPGNDQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSSTLVYVNC 60
Db 277 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYRYTYVDGQKQDFW-CD 335
QY 61 WAAM-----CGGNIRASFGSVNPAFTADTYLQ 88
Db 336 HAAITGSGSYNGITSNVKGTGFKVMSSTNNADTYLE 372

RESULT 11
CIPA_CLOTM
ID CIPA_CLOTM STANDARD; PRT; 1853 AA.
AC Q06851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
DE glycoprotein SI/SL) (Cellulose integrating protein A) (Cohesin).
GN CIPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RX STRAIN=ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RA Demail A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipA) encoding the
RT cellulosomal SI-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RX SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;

```

Tavares G.A., Beguin P., Alzari P.M.;  
"The crystal structure of a type I cohesin domain at 1.7-A resolution.";  
J. Mol. Biol. 273:701-713(1997).  
-!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE CELLULOYTIC ENZYMES.  
-!- SUBCELLULAR LOCATION: Cell surface.  
-!- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC COMPONENTS OF THE CELLULOSE.  
-!- SIMILARITY: Contains 9 cohesin domains.  
-!- SIMILARITY: Contains 2 dockerin domains.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

-----  
EMBL; L08665; -; NOT ANNOTATED\_CDS.  
EMBL; X67506; CAA47840.1; -;  
PIR; S36859; S36859.  
PDB; 1ANU; 23-JUL-97.  
PDB; 1AOH; 08-JUL-98.  
PDB; 1NBC; 26-SEP-97.  
InterPro; IPR001956; CBD 3.  
InterPro; IPR008965; Cellul\_bind.  
InterPro; IPR002102; Cohesin.  
InterPro; IPR002105; Dockerin\_1.  
InterPro; IPR002048; EF-hand.  
Pfam; PF00942; CBM 3; 1.  
Pfam; PF00963; Cohesin; 9.  
Pfam; PF00404; Dockerin\_1; 2.  
ProDom; PD001947; CBD 3; 1.  
PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
PROSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;  
3D-structure.

SIGNAL 1 28  
CHAIN 29 1853 CELLULOSONAL SCAFFOLDING PROTEIN A.  
FT DOMAIN 29 182 COHESIN 1.  
FT DOMAIN 183 322 COHESIN 2.  
FT DOMAIN 323 363 LINKER (PRO/THR-RICH).  
FT DOMAIN 364 522 CELLULOSE-BINDING (BY SIMILARITY).  
FT DOMAIN 523 559 LINKER (PRO/THR-RICH).  
FT DOMAIN 560 704 COHESIN 3.  
FT DOMAIN 724 866 COHESIN 4.  
FT DOMAIN 889 1031 COHESIN 5.  
FT DOMAIN 1054 1196 COHESIN 6.  
FT DOMAIN 1219 1361 COHESIN 7.  
FT DOMAIN 1384 1526 COHESIN 8.  
FT DOMAIN 1548 1690 COHESIN 9.  
FT DOMAIN 1791 1814 DOCKERIN 1.  
FT DOMAIN 1824 1847 DOCKERIN 2.  
FT CONFLICT 1615 1615 A -> AA (IN REF. 1).  
FT STRAND 185 188  
FT STRAND 190 191  
FT TURN 195 196  
FT STRAND 198 206  
FT TURN 210 211  
FT STRAND 213 221  
FT TURN 224 226  
FT STRAND 227 234  
FT TURN 236 237  
FT TURN 243 246  
FT STRAND 247 252  
FT TURN 253 256  
FT STRAND 257 263  
FT TURN 265 266  
FT TURN 270 271  
FT STRAND 273 273

```
FT STRAND 277 286
FT TURN 292 304
FT STRAND 305 306
FT STRAND 309 309
FT STRAND 313 315
FT STRAND 317 319
FT STRAND 369 375
FT STRAND 381 382
FT STRAND 385 385
FT STRAND 387 393
FT STRAND 399 400
FT HELIX 401 403
FT STRAND 404 410
FT STRAND 418 428
FT TURN 430 431
FT STRAND 434 436
FT HELIX 438 440
FT STRAND 441 452
FT TURN 453 454
FT STRAND 455 463
FT STRAND 467 468
FT TURN 470 471
FT STRAND 473 482
FT TURN 483 484
FT STRAND 488 489
FT TURN 491 492
FT TURN 494 495
FT STRAND 498 498
FT STRAND 503 504
FT STRAND 509 512
FT TURN 513 514
FT STRAND 515 518
FT STRAND 1220 1224
FT STRAND 1226 1229
FT TURN 1231 1232
FT STRAND 1234 1242
FT TURN 1246 1247
FT STRAND 1249 1249
FT STRAND 1251 1257
FT TURN 1260 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT HELIX 1279 1282
FT STRAND 1283 1288
FT TURN 1289 1292
FT STRAND 1293 1299
FT TURN 1301 1302
FT TURN 1306 1307
FT STRAND 1309 1309
FT STRAND 1313 1322
FT TURN 1324 1325
FT STRAND 1329 1342
FT TURN 1344 1345
FT STRAND 1348 1348
FT STRAND 1351 1354
FT STRAND 1356 1360
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 35.2%; Score 165; DB 1; Length 1853;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNC 60
Db 365 VSGNLKVEYFNPSDTTNSINPQFVYVNTGSSAIDLSKILRYIYVVGQKDTFW-CD 423
QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88
Db 424 HAAIGSGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

RESULT 12
GUNN_ERWCA
```

```
ID AC Q59394; GUNN_ERWCA STANDARD; PRT; 444 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Atroseptica FCBR C18;
RX MEDLINE=9829944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39788; AAC37033.1; -.
DR HSSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 444;
Best Local Similarity 38.6%; Pred. No. 3.8e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGKVQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSTLVYNC 59
Db 357 TGDVQLQYRVNDDAIRMAVNIKNTGSPKLSDLQVRYVYFHDGKPGANLIV--- 413
QY 60 DWAMCGGNIRASFGSVNPTPTADTYL 87
Db 414 DWANVGPNNTVSTGTTPAASDKNRYV 441

RESULT 13
GUNN_ERWCA STANDARD; PRT; 505 AA.
AC Q47056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
```

```

GN CELV
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: Endoglucanase with some exoglucanase activity.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
CC optimum about 42 degrees Celsius.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76000; CAA53592.1; -.
DR PIR; S39962; S39962.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR ProDom; PD001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 505
FT DOMAIN 32 334
FT DOMAIN 335 352
FT ACT_SITE 168 168
FT ACT_SITE 256 256
FT ACT_SITE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;
SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 505;
Best Local Similarity 38.6%; Pred. No. 4.4e-09;
Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKQVQKNDSAPGDNQIKPGLQVNTGSSVDLSVTYRYWFTRDG--GSSTLVYNC 59
Db 354 TGDVVLQYRNVNDNPSDDAIRMAVNIKGTSTPKLSDLQVRYFPHDCKPGANLFV--- 410

QY 60 DWAMGCGNIRASFGSVNPAATPTADTYL 87
Db 411 DWANVGPNNTVTGTTPAASTDKANRYV 438

RESULT 14
GUNW_ERWCA
ID GUNW_ERWCA STANDARD; PRT; 504 AA.
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
DE (Cellulase V1).

```

```

GN CELV1
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celV1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79241; CAA55823.1; -.
DR PIR; S54744; S54744.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR ProDom; PD001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 504
FT DOMAIN 32 334
FT DOMAIN 335 352
FT ACT_SITE 168 168
FT ACT_SITE 256 256
FT ACT_SITE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;
SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 31.4%; Score 147.5; DB 1; Length 504;
Best Local Similarity 37.5%; Pred. No. 7.4e-09;
Matches 33; Conservative 19; Mismatches 31; Indels 5; Gaps 2;

QY 2 SGGVKQVQKNDSAPGDNQIKPGLQVNTGSSVDLSVTYRYWFTRDG--GSSTLVYNC 59
Db 354 TGDVVLQYRNVNDNPSDDAIRMAFNIKGTSTPKLSDLQVRYFPHDCKPGANLFV--- 410

QY 60 DWAMGCGNIRASFGSVNPAATPTADTYL 87
Db 411 DWANVGPNNTVTGTTPAASTDKANRYV 438

RESULT 15
GUN2_CLOSR
ID GUN2_CLOSR STANDARD; PRT; 914 AA.
AC P50900;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
DE (1,4-beta-cellobiohydrolase II) (Avicelase II).
GN CELY.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 11:45:55 ; Search time 10.2683 Seconds  
(without alignments)  
2704.020 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKNDSAPGDNQ.....IRASFGSVNPTPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	257	54.8	741	16	Q82QF2	Q82qf2 streptomyc
2	228.5	48.7	616	2	Q7X2N2	Q7x2n2 thermomonos
3	225.5	48.1	170	2	Q9RFK6	Q9rfx6 caldibacilli
4	216.5	46.2	930	2	Q9RFK5	Q9rfx5 caldibacilli
5	214.5	45.7	921	2	Q9L8L8	Q9l8l8 caldibacilli
6	209.5	44.7	1091	2	Q8KKF7	Q8kkf7 paenibacilli
7	204.5	43.6	997	2	Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.5	1751	2	Q9AQG4	Q9agg4 caldicellu
9	192.5	41.0	1000	2	Q24820	Q24820 thermophili
10	192.5	41.0	1770	2	Q9X3P5	Q9x3p5 caldicellu
11	191.5	40.8	261	2	Q9AQG7	Q9agg7 caldicellu
12	191.5	40.8	1426	2	Q9X3P6	Q9x3p6 caldicellu
13	191	40.7	1711	2	Q963I1	Q963i1 anaerocellu
14	190.5	40.6	996	2	Q9AQH0	Q9aqh0 caldicellu
15	190.5	40.6	1779	2	Q52374	Q52374 caldicellu
16	190	40.5	499	2	Q93TJ6	Q93tj6 bacillus su

17	190	40.5	508	2	Q93LD0	Q93ld0 bacillus su
18	187	39.9	499	2	Q52731	Q52731 bacillus sp
19	186	39.7	499	2	Q45532	Q45532 bacillus su
20	186	39.7	501	2	Q83012	Q83012 bacillus sp
21	180	38.4	486	2	Q45430	Q45430 bacillus sp
22	177	37.7	499	2	Q8RP06	Q8rp06 bacillus am
23	172.5	36.8	1915	2	Q9RPL0	Q9rp10 acetivibrio
24	168	35.8	473	16	Q9RK75	Q9rk75 streptomyc
25	165.5	35.3	887	2	Q9L3J8	Q9l3j8 clostridium
26	157	33.5	542	2	Q7X3S6	Q7x3s6 bacillus li
27	150.5	32.1	2316	2	Q9FDJ9	Q9fdj9 bacteroides
28	130.5	27.8	1483	16	Q977Y4	Q977y4 clostridium
29	130.5	27.8	1546	2	Q45996	Q45996 clostridium
30	126.5	27.0	1162	2	Q82830	Q82830 clostridium
31	119.5	25.5	307	2	Q46392	Q46392 clostridium
32	117	24.9	1230	2	Q59325	Q59325 clostridium
33	107	22.8	221	2	Q8VVI7	Q8vv17 clostridium
34	87	18.6	440	2	Q93LI9	Q93li9 bacillus sp
35	78	16.6	135	13	Q90542	Q90542 ginglymosto
36	76	16.2	121	13	Q8JGA5	Q8jga5 ginglymosto
37	76	16.2	121	13	Q8JGA9	Q8jga9 ginglymosto
38	75.5	16.1	618	4	Q9Y3Z2	Q9y3z2 homo sapien
39	75.5	16.1	660	2	Q9L3J2	Q9l3j2 clostridium
40	75.5	16.1	986	4	Q94858	Q94858 homo sapien
41	75.5	16.1	1300	4	Q8IZF2	Q8izf2 homo sapien
42	73.5	15.7	547	5	P91006	P91006 caenorhabdi
43	73.5	15.7	1428	16	Q8YRU7	Q8yru7 anabaena sp
44	73	15.6	511	5	Q9GVG5	Q9gyg5 caenorhabdi
45	72	15.4	499	12	Q993M3	Q993m3 autonomic

#### ALIGNMENTS

#### RESULT 1

Q82QF2	PRELIMINARY;	PRT;	741 AA.
ID	Q82QF2		
AC	Q82QF2;		
DT	01-JUN-2003 (TREMELrel. 24, Created)		
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Putative cellulose 1,4-beta-cellobiosidase.		
GN	GUXA1 OR SAV557.		
OS	Streptomyces avermitilis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=33903;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=21477403; PubMed=11572948;		
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,		
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,		
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;		
RT	"Genome sequence of an industrial microorganism Streptomyces		
RT	avermitilis: deducing the ability of producing secondary		
RT	metabolites.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=22608306; PubMed=12692562;		
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,		
RA	Sakaki Y., Hattori M., Omura S.;		
RT	"Complete genome sequence and comparative analysis of the industrial		
RT	microorganism Streptomyces avermitilis.";		
RL	Nat. Biotechnol. 21:526-531(2003).		
DR	EMBL; AP005023; BAC68267.1; -		
DR	GO; GO:0004553; Fhydrylase activity, hydrolyzing O-glycosyl . . . ; IEA.		
DR	GO; GO:0005975; P:carboxylate metabolism; IEA.		
DR	InterPro; IPR001956; CBD 3.		
DR	InterPro; IPR008965; Cellul. bind.		
DR	InterPro; IPR003961; FN_III.		

```
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLYDRLASR6.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Complete proteome.
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B4A7AF CRC64;

Query Match 54.8%; Score 257; DB 16; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.7e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKVQYKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 61
Db 591 SGGLKVLKYNNDSSATDNIAIRGLIVNTGSLDLSKVTRYFTSDGSPTVNAWCY 650
QY 62 AAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 651 AAVGCSNLSKVLVPLTTPVGADAYLE 677

RESULT 2
QYX2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM51;
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tt cel5B, a new endoglucanase encoding
  gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1E1A1A3F5639 CRC64;

Query Match 48.7%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.8e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

QY 2 SGGVKVQYKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 61
Db 471 TGALEVVYRNNSLAADDSQIAPGLRVNTGSLVNTGSLADVEIHYYFTNEPG-GTLQFTCDW 529
QY 62 AAMGCGNIRASFGSVNPTPTADTYL 87
Db 530 AQVGCANVNASFTSL-SAPGADTSL 553

RESULT 3
QYRFK6 PRELIMINARY; PRT; 170 AA.
AC QYRFK6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Caldicellulosum cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosum.
OX NCBI_TaxID=74586;
```

```

[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Summa A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
  Caldicellulosum cellulovorans and action of the recombinant enzyme on
  kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670 (2000).
DR EMBL; AF163837; AAF22273.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F4E3A0B4 CRC64;

Query Match 48.1%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.2e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 3 GGVKQYKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSTLVNCDW 62
Db 20 GSLVQVYRAADTNAGDNQKPKPRIVNRGTSSVPLSELTIYWTVD-GDKPQVFNCDA 78
QY 63 AMGCGNIRASFGSVNPTPTADTYLQ 88
Db 79 QVGCNVRGFSVKLSTGRTGADYIE 104

RESULT 4
QYRFK5 PRELIMINARY; PRT; 930 AA.
AC QYRFK5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN MANA.
OS Caldicellulosum cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosum.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Summa A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
  Caldicellulosum cellulovorans and action of the recombinant enzyme on
  kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670 (2000).
DR EMBL; AF163837; AAF22274.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 46.2%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 6.2e-15;
```



```
Matches 40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;
QY 3 GGVKQVKNDSAPGDNIKPGQLVNTGSSVDLSTVTYRWFTRDGGSTLVVNCWA 62
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 780 GNLVVQYRAADTNQDKPHFRVNRGTSSVPLSELTIRYWTVD-GDKPQVNCWA 838
QY 63 AMGCGNIRASFGSVNPAFTPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 839 QVGCNLRGSLFKLSTGRTGADYIE 864

RESULT 5
Q9L8L8
ID Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Beta-1,4-xylanase Xyna precursor.
GN Xyna.
OS Caldicibacillus cellulosivorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicibacillus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519260; PubMed=11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multidomain 1,4-beta-xylanase from
RT 'Caldicibacillus cellulosivorans' and effect of its xylan-binding domain
RT on enzyme activity.";
RL Microbiology 146:2947-2955(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF200304; AAF61649.1; -.
DR HSSP; Q06851; INCB.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR010000; Glyco_hydro_10.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF02018; CBM 4; 9; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLYDRLASE10.
DR ProDom; PD001947; CBD 3; 2.
DR SMART; SM00633; Glyco_10; 1.
DR Glycosidase; Hydrolase; Signal; Xylan degradation.
KW SIGNAL
FT CHAIN 1 33 POTENTIAL.
SQ SEQUENCE 921 AA; 102380 MW; 65DDDLIAF7567413 CRC64;

Query Match 45.7%; Score 214.5; DB 2; Length 921;
Best Local Similarity 46.5%; Pred. No. 1e-14;
Matches 40; Conservative 16; Mismatches 29; Indels 1; Gaps 1;
QY 3 GGVKQVKNDSAPGDNIKPGQLVNTGSSVDLSTVTYRWFTRDGGSTLVVNCWA 62
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 771 GNLVVQYRAADTNQDKPHFRVNRGTSSVPLSELTIRYWTVD-GDKPQVNCWA 829
QY 63 AMGCGNIRASFGSVNPAFTPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 830 WGCNLRGSLFKLSTGRTGADYIE 855

RESULT 6
Q8KKF7
ID Q8KKF7 PRELIMINARY; PRT; 1091 AA.
AC Q8KKF7;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
```

```
DE Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.1.91).
GN CEL48C.
OS Paenibacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=198119;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RA Sanchez M.M., Pastor F.I.J., Diaz P.;
RT "Paenibacillus sp. BP-23 family 48 cellulase. Cloning and performance
RT on cellulosic substrates.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AJ488933; CAD32945.1; -.
DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00641; fn3; 1.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; PR00844; GLYDRLASE48.
DR ProDom; PD001947; CBD 3; 1.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR SMART; SM00060; FN3; 2.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1091 CELLULOSE 1,4-BETA-CELLOBIOSIDASE.
SQ SEQUENCE 1091 AA; 118001 MW; 21EACCEB2E704478 CRC64;

Query Match 44.7%; Score 209.5; DB 2; Length 1091;
Best Local Similarity 46.0%; Pred. No. 4.4e-14;
Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY 2 SGGVKQVKNDSAPGDNIKPGQLVNTGSSVDLSTVTYRWFTRDGGSTLVVNCWD 61
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 940 TGTLEQVYRNGSSASGNATPQFNKNUTGTTADLSKVKRYFTKD-SAADMSFWCDY 998
QY 62 AAMCGNIRASFGSVNPAFTPTADTYLQ 88
   :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 999 AQLGSANVQGSFVAVNPAKGTADTYLE 1025

RESULT 7
Q9Z4I1
ID Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
```

```
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR001956; CBD 3.
DR InterPro: IPR008965; Cellul bind.
DR InterPro: IPR003961; FN III-like.
DR InterPro: IPR008957; FN III-like.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR008928; Glyco_trans_6hp.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR Pfam: PF00942; CBM 3; 2.
DR Pfam: PF00041; fn3_2.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR ProDom: PD001947; CBD 3; 1.
DR SMART: SM00060; FN3; 2.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 POTENTIAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.6%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.4e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSDAPGDNQIKFGLQVNTGSSVDLSTVTYVWFTTRDGGSTLVNCDW 61
Db 846 TGTLEQVYRSGSGNSNAVTQFNLKNTGTQAIIDLSTVKIRYFTKO-GTELSFWCDY 904

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 905 AQVGSANVQGMFVAVNPAKGTADTYVE 931

RESULT 8
Q9AQG4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQG4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CelB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; Q06851; INBC.
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR001956; CBD 3.
DR InterPro: IPR008965; Cellul bind.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR008928; Glyco_trans_6hp.
DR InterPro: IPR000566; Lipocin_cytFABP.
DR Pfam: PF00942; CBM 3; 4.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR ProDom: PD001947; CBD 3; 3.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSDAPGDNQIKFGLQVNTGSSVDLSTVTYVWFTTRDGGSTLVNCDW 61
Db 372 SGQIKVLYANKETNTNTIRPWLKVNTGSSSIDLSRVIRYVWYTVVDGDKAQSAVS-DW 430

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 431 AQIGASNVTFKFKVLSSSVSGADYILE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
ID Q9X3P5;
SQ SEQUENCE 1770 AA; 192176 MW; 60178CBF3C00BE95 CRC64;
```

```
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE KYNA.
GN XYN.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF078737; AAD30363.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR006584; CBD IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBM 3; 3.
DR Pfam; PF02018; CBM 4; 9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD 3; 3.
DR SMART; SM00606; CBD IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPCLQIVNTGSSVDLSVTYVRYWFTRDGSSSTLVNCDWAA 63
Db 1104 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPOSAB-CDWAA 1162
QY 64 MCGNIRASFGSVNTPATPTADTYLQ 88
Db 1163 IGASNVTFNFKLVSSGVSGADYYLE 1187

RESULT 11
Q9A0G7 PRELIMINARY; PRT; 261 AA.
AC Q9A0G7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 6 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;

Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
Bergquist P.L.;
"Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
Curr. Microbiol. 40:333-340(2000).
-!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
EMBL; AF078737; AAD30364.1; -.
HSSP; Q06851; INBC.
GO; GO:0003779; F:actin binding; IEA.
GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
GO; GO:0005975; P:carbohydrate metabolism; IEA.
InterPro; IPR001589; Actbind_actnin.
InterPro; IPR001956; CBD 3.
InterPro; IPR008965; Cellul_bind.
InterPro; IPR001000; Glyco_hydro_10.
InterPro; IPR001547; Glyco_hydro_5.
Pfam; PF00942; CBM 3; 3.
Pfam; PF00150; cellulase; 1.
Pfam; PF00331; Glyco_hydro_10; 1.
PRINTS; PR00134; GLHYDRLASE10.
ProDom; PD001947; CBD 3; 3.
SMART; SM00633; Glyco_10; 1.
PROSITE; PS00019; ACTININ 1; 1.
PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
```

```
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 261 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.3e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPCLQIVNTGSSVDLSVTYVRYWFTRDGSSSTLVNCDWAA 63
Db 118 GLKVLKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPOSAB-CDWAA 176
QY 64 MCGNIRASFGSVNTPATPTADTYLQ 88
Db 177 IGASNVTFNFKLVSSGVSGADYYLE 201

RESULT 12
Q9X3P6 PRELIMINARY; PRT; 1426 AA.
ID Q9X3P6;
AC Q9X3P6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CelB.
GN CELB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RL thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
Pfam; PF00942; CBM 3; 3.
Pfam; PF00150; cellulase; 1.
Pfam; PF00331; Glyco_hydro_10; 1.
PRINTS; PR00134; GLHYDRLASE10.
ProDom; PD001947; CBD_3; 3.
SMART; SM00633; Glyco_10; 1.
PROSITE; PS00019; ACTININ 1; 1.
PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
```

```
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 1426;
Best Local Similarity 43.5%; Pred. No. 5.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAA 63
Db 413 GLKVLVKNNETSASTGSIKPFVKVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAAQ 471

QY 64 MCGGNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGVADYILE 496

RESULT 13
ID P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9491383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme
RT (Cela) from the extreme thermophile Anaerocellum thermophilum with
RT separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P26221; 1TF4.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHDLRLASE48.
DR ProDom; PD001947; CBD 3; 2.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
FT NON TER 1
FT SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.7%; Score 191; DB 2; Length 1711;
Best Local Similarity 43.8%; Pred. No. 8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-KVVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAA 59
Db 683 VAGGQIKVLVYANKETSTNTIRPMLKVNTGSSVDLSRVKIRYWTVDGDKPOS AV-741

QY 60 DWAMCGGNIRASFGSVNPTPTADTYLQ 88
Db 742 DWAIQIGASNVTFKFKVLSGSGVADYILE 770

RESULT 14
ID Q9AQHO PRELIMINARY; PRT; 996 AA.
AC Q9AQHO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON TER 996
FT SEQUENCE 996 AA; 108275 MW; 3C72B6ED22E3C614 CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.8e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAA 63
Db 844 GLKVLVKNNETSASTGSIKPFVKVNGSSVDLSRVKIRYWTVDGDKPOS AV-CDWAAQ 902

QY 64 MCGGNIRASFGSVNPTPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGVADYILE 927

RESULT 15
ID O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
```

Search completed: May 11, 2004, 12:09:29  
Job time : 11.2683 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:04:16 ; Search time 4.31843 Seconds  
(without alignments)  
1052.023 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKXNDAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pap.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pap.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pap.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pap.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pap.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.7	616	3	US-09-136-574A-47
2	194.5	41.5	1751	3	US-09-136-574A-44
3	191.5	40.8	1426	3	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	476	4	US-09-339-159B-4
8	165	35.2	493	3	US-09-198-956-10
9	165	35.2	493	3	US-09-198-955A-12
10	165	35.2	493	4	US-09-694-531-12
11	165	35.2	493	4	US-09-670-141-10
12	165	35.2	493	4	US-10-072-152-12
13	120.5	25.7	531	2	US-07-862-588B-7
14	114	24.3	162	1	US-08-048-164A-2
15	114	24.3	162	1	US-08-460-462-2
16	114	24.3	162	1	US-08-460-457-2
17	114	24.3	162	1	US-08-460-458-2
18	114	24.3	162	1	US-08-460-455-2
19	114	24.3	162	2	US-08-330-394A-2
20	114	24.3	163	3	US-09-006-636-7
21	114	24.3	163	3	US-09-006-632-7
22	114	24.3	163	4	US-09-325-274-7
23	113	24.1	382	3	US-09-277-716-22
24	113	24.1	382	4	US-09-609-161B-22
25	112	23.9	154	2	US-08-330-394A-29
26	112	23.9	156	2	US-08-330-394A-22
27	64	13.6	143	4	US-09-301-593-26

28	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appli
29	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
30	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
31	63.5	13.5	1785	4	US-09-341-587-3	Sequence 3, Appli
32	63	13.4	453	4	US-09-301-593-18	Sequence 18, Appl
33	63	13.4	472	4	US-09-301-593-43	Sequence 43, Appl
34	63	13.4	718	4	US-09-328-352-4640	Sequence 4640, Ap
35	63	13.4	1581	3	US-09-110-517-2	Sequence 2, Appli
36	62.5	13.3	288	4	US-09-423-439-38	Sequence 38, Appl
37	62.5	13.3	445	1	US-08-353-400-33	Sequence 33, Appl
38	62.5	13.3	641	4	US-09-687-538B-8	Sequence 8, Appli
39	62.5	13.3	673	4	US-09-423-439-32	Sequence 32, Appl
40	62.5	13.3	802	3	US-09-081-345-18	Sequence 18, Appl
41	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appli
42	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
43	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
44	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
45	61.5	13.1	307	3	US-08-370-476-68	Sequence 68, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 616 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 3; Length 616;  
Best Local Similarity 43.2%; Pred. No. 2.4e-13;  
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;  
QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSSTVTVYVWFTRDGGSSTLVYNC 60  
DB 1 MGSQVLYKNNETSASTGSRPWFVKLVNGSSVDLSRVKIRYWTVDGDKPOSQAV-CD 59  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 60 WAQIGASNTFNFVKLSGSGVADYYLE 87

RESULT 2

US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1751 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44  
Query Match 41.5%; Score 194.5; DB 3; Length 1751;  
Best Local Similarity 44.7%; Pred. No. 1.1e-12;  
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;  
QY 4 GYKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSSTVTVYVWFTRDGGSSTLVYNC 63  
DB 4 GYKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSSTVTVYVWFTRDGGSSTLVYNC 63

DB 678 GVKVLYKNNETSASTGSRPWFVKLVNGSSVDLSRVKIRYWTVDGDKPOSQAV-CD 736

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88  
DB 737 IGASNTFNFVKLSGSGVADYYLE 761

RESULT 3

US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1426 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6294366e  
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43  
Query Match 40.8%; Score 191.5; DB 3; Length 1426;  
Best Local Similarity 43.5%; Pred. No. 1.8e-12;  
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;  
QY 4 GYKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSSTVTVYVWFTRDGGSSTLVYNC 63  
DB 413 GLKVLKNNETSASTGSRPWFVKLVNGSSVDLSRVKIRYWTVDGDKPOSQAV-CD 471  
QY 64 MCGNIRASFGSVNPATPTADTYLQ 88  
DB 472 IGASNTFNFVKLSGSGVADYYLE 496

RESULT 4

```

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.1e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKQYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNC 59
DB 549 VNSDLVQYKQDNRNATDNQIKPFNTQNGTSPVLSLTLRYFTK-DSSAAMNGWI 607
QY 60 DWAAMGCGNIRASFGSNVPATPTADTYLQ 88
DB 608 DWAKLGSNIQISFGNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Shlein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59580830 No. 5958083disk of No. 5958083th America, Inc.

```

```

; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 35.3%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 4.1e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVQYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNC 61
DB 402 TGNLVVQYKVGDTSATDNQMKPSFNKNNGTTPVNLGSLKRLRYFTKD-GTADMSASPDW 460
QY 62 AAMGCGNIRASFGSNVPATPTADTYLQ 88
DB 461 AQIGASNVSAAP--ANFTGSNTDTYVE 485

RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brody and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
;
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.2%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 1.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSTLYVNC 60
DB 5 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLKLTLYRYTYVDGQKDTFW-CD 63
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
DB 64 HAAIGSGSYNGITSNVKGT FVKMSSTNNADTYLE 100

RESULT 7
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.2%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSTLYVNC 60
DB 314 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLKLTLYRYTYVDGQKDTFW-CD 372
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
DB 373 HAAIGSGSYNGITSNVKGT FVKMSSTNNADTYLE 409

RESULT 8
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.2%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 3.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSTLYVNC 60
DB 314 VSGNLKVEFYNSPDSITNSINPQFKVTNTGSSAIDLKLTLYRYTYVDGQKDTFW-CD 372
QY 61 WAAM-----GCNIRASFGSVNPATPTADTYLQ 88
DB 373 HAAIGSGSYNGITSNVKGT FVKMSSTNNADTYLE 409

RESULT 9
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Schulein, Lene E.
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
```

```
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match          35.2%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 331 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLKSLRLRYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426

RESULT 10
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; CURRENT FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match          35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 331 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLKSLRLRYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426

RESULT 11
US-09-670-141-10
; Sequence 10, Application US/09670141
```

```
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match          35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 4.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 331 VSGNLKVEFYNSPDTTNSINPQFKVTNTGSSAIDLKSLRLRYYYTVDGKQDTFW-CD 389

QY 61 WAAM-----CCGNIRASFGSVNPATPTADTYLQ 88
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 390 HAAIIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426

RESULT 12
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
```

SEQ ID NO 12  
LENGTH: 493  
TYPE: PRT  
ORGANISM: Clostridium thermocellum  
US-10-072-152-12

Query Match 35.2%; Score 165; DB 4; Length 493;  
Best Local Similarity 36.1%; Pred. No. 4.1e-10;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGVKVQYKNDSPAGDQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCW 60  
DB 331 VSGNLKVEFYNSPDDTNSINPQKVTNGSSADLSKLTURLYTYVVDGQKQTFW-CD 389

QY 61 WAAM-----CGCNRASFGSVNPATPTADTYLQ 88  
DB 330 HAAIGSNGSYNGITSNVKGFVKSSSTNNADTYL 426

RESULT 13  
US-07-862-588B-7  
Sequence 7, Application US/07862588B  
Patent No. 5916796  
GENERAL INFORMATION:  
APPLICANT: Joergensen, Per Linea  
APPLICANT: Sch lein, Martin  
APPLICANT: Hansen, Christian  
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
STREET: 405 Lexington Avenue, 62nd floor  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10017

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,588B  
FILING DATE: 19920727  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 164/90  
FILING DATE: 19-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00013  
FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Zelson, Steve T. / Lambiris, Elias J.  
REGISTRATION NUMBER: 30,335 / 33,728  
REFERENCE/DOCKET NUMBER: 3425.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 531 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-862-588B-7

Query Match 25.7%; Score 120.5; DB 2; Length 531;  
Best Local Similarity 32.9%; Pred. No. 3.5e-05;  
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVKVQYKNDSPAGDQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCW 61  
DB 432 TGNLVQYKVGDTSATDNQKPSFNKNGTTPVNLGSLKXXXXKD-GPADMSCSIDW 490

QY 62 AAMGCGNIRASFGSVNPATPTADTY 86  
DB 491 AQLGRTNVLLAF--ANFTGSNTDTY 513

RESULT 14  
US-08-048-164A-2  
Sequence 2, Application US/08048164A  
Patent No. 5496934  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Doi, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/048,164A  
FILING DATE: 14-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-048-164A-2

Query Match 24.3%; Score 114; DB 1; Length 162;  
Best Local Similarity 29.9%; Pred. No. 4.2e-05;  
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY 2 SGGVKVQYKNDSPAGDQIKPGLQLVNTGSSVDLSTVTYVYFTRDGGSSTLVYNCW 61  
DB 3 TSSMSVEFYNSKSAQTSITPIKITNTSDNLNDVKRYVYTSDTGQTGQTFW-CDH 61

QY 62 AAMGCGN-----IRASF--GSVNPATPTADTYLQ 88  
DB 62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97

RESULT 15  
US-08-460-462-2  
Sequence 2, Application US/08460462  
Patent No. 5670623  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Doi, Roy H.  
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS  
NUMBER OF SEQUENCES: 21

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

```

```

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 4.2e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

Qy      2 SGGVKVQYKNDSPGDNQIKPGQLVNTGSSVDLSITVYVWFTRDGSGSTLYVNCDW 61
Db      3 TSSMSVEFYNSKSAQTNSITPIIKITNTSDSLNLDNDVKRYYYTSDGTQGTFW-CDH 61

Qy      62 AAMGCGN-----IRASF--GSVNPATPTADTVLQ 88
Db      62 AGALLGNSYDNTSKVTANFVKETASP-TSTYDTIVE 97

```

Search completed: May 11, 2004, 12:11:22  
Job time : 5.31843 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 12:09:36 ; Search time 10.2693 Seconds  
(without alignments)  
2378.773 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 459

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	88	12	US-09-917-376-5
2	469	100.0	88	14	US-10-155-400-5
3	469	100.0	89	12	US-09-917-376-4
4	469	100.0	89	14	US-10-155-400-4
5	469	100.0	154	10	US-09-917-378-4
6	469	100.0	762	10	US-09-917-378-1
7	466	99.4	150	10	US-09-917-384-5
8	466	99.4	150	10	US-09-917-383-5
9	466	99.4	1043	10	US-09-917-384-6
10	466	99.4	1043	10	US-09-917-383-6
11	466	99.4	1228	10	US-09-917-384-1
12	466	99.4	1228	10	US-09-917-383-1
13	462	98.5	957	12	US-09-917-376-1
14	462	98.5	957	14	US-10-155-400-1
15	257	54.8	741	14	US-10-156-761-8100

16	190	40.5	508	15	US-10-369-493-23151	Sequence 23151, A
17	180	38.4	1621	14	US-10-185-990-10	Sequence 10, Appl
18	165	35.2	476	12	US-10-372-054-4	Sequence 4, Appl
19	165	35.2	493	12	US-10-655-433-12	Sequence 12, Appl
20	165	35.2	493	13	US-10-072-152-12	Sequence 12, Appl
21	165	35.2	599	10	US-09-955-555A-29	Sequence 29, Appl
22	155.5	33.2	1352	10	US-09-784-554B-2	Sequence 2, Appl
23	150.5	32.1	1350	10	US-09-784-554B-4	Sequence 4, Appl
24	130.5	27.8	1483	12	US-10-282-122A-51483	Sequence 6, Appl
25	114	24.3	256	14	US-10-261-446-6	Sequence 51483, A
26	113	24.1	382	10	US-09-808-898-22	Sequence 22, Appl
27	105.5	22.5	163	12	US-10-460-524-2	Sequence 2, Appl
28	78.5	16.7	1049	12	US-10-282-122A-49900	Sequence 49900, A
29	75.5	16.1	618	12	US-10-211-462-223	Sequence 223, App
30	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appl
31	75.5	16.1	986	12	US-10-312-312-53	Sequence 53, Appl
32	75.5	16.1	986	14	US-10-120-604-101	Sequence 101, App
33	75.5	16.1	986	14	US-10-225-567A-406	Sequence 406, App
34	75.5	16.1	986	16	US-10-398-458-3	Sequence 3, Appl
35	75.5	16.1	1131	16	US-10-398-458-2	Sequence 2, Appl
36	75.5	16.1	1325	9	US-09-747-835A-24	Sequence 24, Appl
37	75.5	16.1	1325	12	US-10-312-312-24	Sequence 24, Appl
38	75.5	16.1	1346	9	US-09-747-835A-20	Sequence 20, Appl
39	75.5	16.1	1346	9	US-09-747-835A-61	Sequence 61, Appl
40	75.5	16.1	1346	12	US-10-312-312-20	Sequence 20, Appl
41	75.5	16.1	1346	12	US-10-312-312-61	Sequence 61, Appl
42	75.5	16.1	1371	9	US-09-747-835A-25	Sequence 25, Appl
43	75.5	16.1	1371	12	US-10-312-312-25	Sequence 25, Appl
44	75	16.0	161	12	US-10-424-599-196885	Sequence 196885, App
45	72.5	15.5	69	9	US-09-764-860-377	Sequence 377, App

ALIGNMENTS

RESULT 1

US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain

Query Match 100.0%; Score 469; DB 12; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.5e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYWFTRDGGSTLVNCD	60
Db	1	VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYWFTRDGGSTLVNCD	60
QY	61	WAAMCGNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMCGNIRASFGSVNPATPTADTYLQ	88

RESULT 2

US-10-155-400-5  
; Sequence 5, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-5

Query Match 100.0%; Score 469; DB 14; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.5e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||  
DB 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||  
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||

RESULT 3  
US-09-917-376-4  
; Sequence 4, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-4

Query Match 100.0%; Score 469; DB 12; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.5e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||  
DB 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||

US-10-155-400-4  
; Sequence 4, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-4

Query Match 100.0%; Score 469; DB 14; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.5e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||  
DB 1 VSGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYNCD 60  
|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||  
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||

RESULT 5  
US-09-917-378-4  
; Sequence 4, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate  
; OTHER INFORMATION: binding domain  
US-09-917-378-4

Query Match 100.0%; Score 469; DB 10; Length 154;  
Best Local Similarity 100.0%; Pred. No. 2.9e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
Db 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||  
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||

## RESULT 6

US-09-917-378-1  
; Sequence 1, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-1

Query Match 100.0%; Score 469; DB 10; Length 762;  
Best Local Similarity 100.0%; Pred. No. 2e-47;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
Db 455 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 514  
|||||  
QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||  
Db 515 WAAMCGCNIRASFGSVNPATPTADTYLQ 542  
|||||

## RESULT 7

US-09-917-384-5  
; Sequence 5, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-5

Query Match 99.4%; Score 466; DB 10; Length 150;

Best Local Similarity 98.9%; Pred. No. 6.5e-48;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
Db 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||  
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||

## RESULT 8

US-09-917-383-5  
; Sequence 5, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-5

Query Match 99.4%; Score 466; DB 10; Length 150;  
Best Local Similarity 98.9%; Pred. No. 6.5e-48;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
Db 1 VSGGVKQVQKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTTRDGSSTLVNCD 60  
|||||  
QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||  
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88  
|||||

## RESULT 9

US-09-917-384-6  
; Sequence 6, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

```

; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: GuxA
US-09-917-384-6

```

Query Match	99.4%	Score 466;	DB 10;	Length 1043;
Best Local Similarity	98.9%	Pred. NO. 6.8e-47;		
Matches 87;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;

61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
|||  
537 WAAMCGGNIRASFGSVNPATPTADTYLQ 564

RESULT 10  
 US-09-917-383-6  
 ; Sequence 6, Application US/09917383  
 ; Publication No. US20030104522A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,383  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 1043  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment 6  
 ; OTHER INFORMATION: Guxa  
 ; US-09-917-383-6

```
Query Match      99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred.No. 6.8e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
|||||  
537 WAAMCGGNIRASFGSVNPATPTADTYLQ 564

RESULT 11  
 US-09-917-384-1  
 ; Sequence 1, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYLICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1

```

: SEQ ID NO 1
:
: LENGTH: 1228
:
: TYPE: PRT
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: Description of Artificial Sequence: Segment of
:
: OTHER INFORMATION: GuxA
:
: US-09-917-394-1

```

```
Query Match      99.4%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
|||  
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLO 671

```

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication NO. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment 1
; OTHER INFORMATION: GuxA
US-09-917-383-1

```

```
Query Match      99.4%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 8.3e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 51 WAAMGCNIRASFGSVNPATPTADTYLQ 88  
 |||||  
 644 WAAMGCNIRASFGSVNPATPTADTYLO 671  
 |||||

RESULT 13  
US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOOLYTICUS  
; FILE REFERENCE: 40197.4US01



; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 98.5%; Score 462; DB 12; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 60  
DB 869 VSGGVKQYKNDSPAGDNQIKPGLQVVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14  
US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE OF INVENTION: CELLULOXYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 98.5%; Score 462; DB 14; Length 957;  
Best Local Similarity 97.7%; Pred. No. 1.9e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 60  
DB 869 VSGGVKQYKNDSPAGDNQIKPGLQVVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15  
US-10-156-761-8100  
; Sequence 8100, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8100  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100  
Query Match 54.8%; Score 257; DB 14; Length 741;  
Best Local Similarity 56.3%; Pred. No. 6e-22;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
QY 2 SGGVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 61  
DB 591 SGGKLVYKNDSSATDNARPGIRIVNTGSGSLDLSKVTARYYFTRDGSGSPTVNAWCDY 650  
QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 651 AAVGCSNVSLKVVPLTTPVPGADAYLE 677  
Search completed: May 11, 2004, 12:25:48  
Job time : 10.2883 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:40:05 ; Search time 1264.92 Seconds  
(without alignments)  
3015.368 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKVQYKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO-epool/US09917376/runat\_11052004\_114535\_28241/app\_query.fasta\_1.1429  
-DB=GenEmbl -QWTFastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CGN 1 1 7742 @runat\_11052004\_114535\_28241 -NCPUL=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pln:\*  
35: em\_hg\_rod:\*  
36: em\_hg\_mam:\*  
37: em\_hg\_vrt:\*  
38: em\_sv:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	469	100.0	2289	6	AX700050	Sequence	
2	466	99.4	3687	6	AX700036	Sequence	
3	462	98.5	2869	6	AX700058	Sequence	
4	436	93.0	3365	6	AX700025	Sequence	
5	257	54.8	299175	1	AP005023	Streptomyces	
6	228.5	48.7	1957	1	AY298814	Thermophilus	
7	225.5	48.1	4567	1	AF163837	Caldibacillus	
8	214.5	45.7	3237	1	AF200304	Caldibacillus	
9	209.5	44.7	3509	1	PAE488933	Paenibacillus	
10	204.5	43.6	4161	1	BSP133614	Bacillus	
11	195.5	41.7	2029	6	E35142	Truncated clone	
12	195.5	41.7	2029	6	E35143	Truncated clone	
13	194.5	41.5	6005	1	AF078038S5	Caldicellum	
14	194.5	41.5	6416	6	E35100	Truncated clone	
15	192.5	41.0	4743	1	AB008029	Thermophilus	
16	192.5	41.0	11707	1	AF078737	Caldicellum	
17	192.5	41.0	11707	6	E35099	Truncated clone	
18	191.5	40.8	787	1	AF078038S3	Caldicellum	
19	191	40.7	5513	1	ATZ86105	A. thermophilus	
20	190.5	40.6	3262	1	AF078038S1	Caldicellum	
21	190.5	40.6	5437	1	CASR69XYN2	Caldicellum	
22	190	40.5	1523	1	AF355629	Bacillus	
23	190	40.5	1800	1	BSEGLUG	Bacillus subtilis	
24	190	40.5	1920	1	BACGLUB	B. subtilis	
25	190	40.5	2314	1	BS14GLUC	B. subtilis	
26	190	40.5	2435	1	BSBGLUC2	Bacillus subtilis	
27	190	40.5	2589	1	AY044252	Bacillus subtilis	
28	190	40.5	26170	1	BC170DEGR	B. subtilis	
29	190	40.5	207829	1	BSUB0010	Bacillus subtilis	
30	188	40.1	1928	1	BACMCASE	B. subtilis	
31	187	39.9	1593	1	AY183475	Bacillus subtilis	
32	187	39.9	2084	1	AF045482	Bacillus subtilis	
33	186.5	39.8	4977	1	CDCWANA	Caldococcus	
34	186.5	39.8	5439	1	CDCELA	Caldococcus	
35	186	39.7	2175	1	BACELD	B. subtilis	
36	186	39.7	2175	6	E05425	DNA sequencing	
37	186	39.7	2587	1	AB016164	Bacillus	
38	182.5	38.9	2977	6	A28170	B. lautus strain	
39	182.5	38.9	4241	1	CSCELB	Caldococcus	
40	182.5	38.9	5284	1	CSU16308	Caldococcus	
41	180.5	38.5	2831	1	BACELAA	Bacillus lautus	
42	180	38.4	1553	1	BSU27084	Bacillus sp.	
43	177	37.7	2152	1	AF363635	Bacillus	
44	172.5	36.8	6555	1	AF155197	Acetivibrio	
45	168	35.8	291000	1	SCO939105	Streptomyces	

#### ALIGNMENTS

RESULT 1

```

AX700050
LOCUS AX700050 2289 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012110.
ACCESSION AX700050
VERSION AX700050.1 GI:29536020
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant mannanase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012109-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..2289
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,21e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-5 (1-88) x AX700050 (1-2289)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATACCA 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGTCTCCAGTTGGTGAATACGGGTCTGTCGGTGGATTGTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGTTCGGTACTGGTTCACCCGGGATGGTGGTCTGTCGACTGGTGTACAACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGTGTGGGAATATCCGGGCTCGTTCGGTTCGGTGAACCCGGCGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B., Himmel, M.E. and Decker, S.R.
TITLE Thermal tolerant cellulase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012109-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of GuxA"

ORIGIN
Alignment Scores:
Pred. No.: 2,21e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-5 (1-88) x AX700050 (1-2289)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATACCA 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGTCTCCAGTTGGTGAATACGGGTCTGTCGGTGGATTGTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGTTCGGTACTGGTTCACCCGGGATGGTGGTCTGTCGACTGGTGTACAACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGTGTGGGAATATCCGGGCTCGTTCGGTTCGGTGAACCCGGCGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE Acidothermus cellulolyticus
ORGANISM Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant avicelase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012090-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,19e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-5 (1-88) x AX700058 (1-2869)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAAGAAATATGATTCGGCGCGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTTTCAGGTGGTGAATACCGGGTCTGTCGGTGGATTGTCGACGGTG 2724

```

```

ORIGIN
Alignment Scores:
Pred. No.: 9,53e-48 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-5 (1-88) x AX700036 (1-3687)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATAACCA 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTGTCGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGTTCGGTACTGGTTCACCCGGGATGGTGGTCTGTCGACTGGTGTCAACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGTGTGGGAATATCCGGGCTCGTTCGGGCTCGTGAACCCGGCGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013

RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE Acidothermus cellulolyticus
ORGANISM Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant avicelase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012090-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1..2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 2,19e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-5 (1-88) x AX700058 (1-2869)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAAGAAATATGATTCGGCGCGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTTTCAGGTGGTGAATACCGGGTCTGTCGGTGGATTGTCGACGGTG 2724

```

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 2725 ACGGTGGGGTACTGGTTCACCCGGGATGGTGGCTCGTCGACACTGGGTGTACAACTGTGAC 2784

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleAraGlySerPheGlySerValAsnProAlaThr 80  
 Db 2785 TGGGCGGCGATCGGGTGTGGGAATATCCGCGCCTCGTTCGGCTCGGTGAACCCGGCGACG 2844

Qy 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 2845 CCGACGGCGGACCACTACCTGCAG 2868

RESULT 4  
 AX700025 3365 bp DNA linear PAT 03-APR-2003  
 LOCUS Sequence 2 from Patent WO03012095.  
 DEFINITION AX700025  
 ACCESSION AX700025  
 VERSION AX700025.1 GI:29536018  
 KEYWORDS  
 SOURCE Acidothermus cellulolyticus  
 ORGANISM Acidothermus cellulolyticus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Frankineae; Acidothermaceae; Acidothermus.

REFERENCE  
 1 Adney,W.S., Ding,S.Y., Vinzant,T.B., Himmel,M.E., Decker,S.R. and Lantz McCarter,S.  
 Thermal tolerant exoglucanase from Acidothermus cellulolyticus  
 Patent: WO 03012095-A 2 13-FEB-2003;  
 JOURNAL Midwest Research Institute (US)  
 FEATURES  
 source Location/Qualifiers  
 1..3365  
 /organism="Acidothermus cellulolyticus"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:28049"

ORIGIN

Alignment Scores:  
 Pred. No.: 4,72e-44 Length: 3365  
 Score: 436.00 Matches: 81  
 Percent Similarity: 97.62% Conservative: 1  
 Best Local Similarity: 96.43% Mismatches: 2  
 Query Match: 92.96% Indels: 0  
 DB: Gaps: 0

US-09-917-376-5 (1-88) x AX700025 (1-3365)

Qy 5 ValIysValGlnTyrIlysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 112 CTCAAAGCGGAGPATAGAAACAATGATTCGGCGCGAGTGACCAACAGATCAAAACCGGGT 171

Qy 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44  
 Db 172 CTCAGTTGGTGAATACCGGGTGTCTCGGTGGATTTCGACGGTGACGGTGGCGTAC 231

Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
 Db 232 TGGTTACCCCGGATGGTGGTCTCGACACTGGTGTACAACCTGACTGGCGCGCATG 291

Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
 Db 292 GGGTGTGGGAATATCCGCGCCTCTGTTGGCTCGGTGAACCCGGCGACGCGCGCGAC 351

Qy 85 ThrTyrLeuGln 88  
 Db 352 ACCTACCTGCAG 363

RESULT 5  
 AP005023 299175 bp DNA linear BCT 10-MAY-2003  
 LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 3/30.  
 DEFINITION AP005023  
 ACCESSION AP005023.1 BA000030  
 VERSION AP005023.1 GI:29604083

Streptomyces avermitilis MA-4680  
Streptomyces avermitilis MA-4680  
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.

1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.  
Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
11572948

2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,  
Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.  
Complete genome sequence and comparative analysis of the industrial  
microorganism Streptomyces avermitilis  
Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306

3 (bases 1 to 299175)  
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,  
Sakaki, Y. and Hattori, M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center,  
National Institute of Technology and Evaluation, Biotechnology  
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nite.go.jp, URL: <http://www.bio.nite.go.jp/>,  
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun  
Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi  
Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi  
Osonoe(\*4), Norihiro Kushida(\*4), Hisashi Kikuchi(\*4), Tadayoshi  
Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7)  
and Satoshi Omura(\*1,\*3)  
Final finishing process and all annotation were done by H. Ikeda  
and J. Ishikawa.

\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
<http://avermitilis.ls.kitasato-u.ac.jp>.

Location/Qualifiers  
1. .299175  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267,  
NCIMB 12904 or NRRL 8165.-synonym: Streptomyces  
avermectinius"  
complement(155. .550)  
/note="SAV445  
Tn4561, truncated"  
/codon\_start=1  
/transl\_table=11  
/product="putative transposase"  
/protein\_id="BAC68155.1"  
/db\_xref="GI:29604084"  
/translation="WAGPPRSRTTSRTTADREDSAHPADPEPGVRRIOKVQANLOV  
EGHHALARRAGQLYQRYPQGMEDQIGALGLVNLVLFNTRYMDALTQLRADGFVR  
DSDVRLSPFMRHNVLRGFSQLPDL"  
complement(474. .740)  
/note="IS630 family, truncated"

```
SAV446"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604085"
/translation="MPDIONAGPSTSVTAAAVSTAGSVYMMAPTSFVMAAPPSPGRRR
EPGSLGAVRAYDVIRMLSRDGRPTPLGDAITYGGSRLCTSCA"
766..987
/note="SAV447"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68157.1"
/db_xref="GI:29604086"
/translation="MVVAVGCAITSVPTAGGVAAEPVCVRRLTDGQKLQIVRRG
SASSVFRFRAMLLASAGNRVPVIAQLV"
complement(1954..2571)
/note="SAV448"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68158.1"
/db_xref="GI:29604087"
/translation="MPTSPRARERANTRERIEAALHVLTEGIAALTRIRIATDVE
YSAPVYQHFANKDALVLEVAHGHRMLSEFSQAQEPDTRRMTRIASQVYRFAGE
HPHLQVMDPMVDADERRVAEPVIDLKELLTWSAAHDVVLADFDQACEILWCTL
YGIASLYLGSVGNDRARYLAEQALHTILLGWRTEPANGRPASS"
2694..3098
/note="SAV449"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68159.1"
/db_xref="GI:29604088"
/translation="MIYHNIRVTMKATATPEQIEGVLESWRNGRSPNAIKGFVWGRN
HGDDYGAVFVEHLDGLFAYLTHPTTYQTDHLGLHLVERLEIFDVSDDDDPLNAK
IQELHRRRNEINPQIAGMLADVPTYTSGVDD"
3127..4029
/note="SAV450"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68160.1"
/db_xref="GI:29604089"
/translation="NAKHVILIGASIGTNNVARLLAEGSGSVIRVTRSGSGPEHPLVDR
VAADSDRLTELSRGAGVITYCANPPSVTWERLLPLOTAAITAANKANDAVLALT
GSLYAGPQPGGRMNEHTPMAATGHKRLRRRMEQALAGGIRTVVEVRGSDYIGKAN
GIYSLILKALQKGAAMITGHLDMPHTFTFNGDMAQALITLASEERAWGRPHWVPSA
PAVTIRELARYAAGRPPIKLIQIPRSVTRTAGLIVPIAREMAEMDYQWYAPFMD
ATETTETGLTATDLDTAVRNEVG"
complement(4130..4498)
/gene="bidg"
complement(4130..4498)
/gene="bidg"
/note="SAV451"
/codon_start=1
/transl_table=11
/product="putative anti-anti-sigma factor"
/protein_id="BAC68161.1"
/db_xref="GI:29604090"
/translation="WTDAPHIDVPHDAAVIALSGEFDIAAAPAVARALELIANG
HPDLVADLSGVTFCDSSGALVGIWRCAKDAGSLTLAAIPDRLSLLSVTGMDAPL
PAYSADAAALAAQGNRTTA"
4635..5156
/note="SAV452"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68162.1"
/db_xref="GI:29604091"
/translation="MFAVKESVRVVRVGEIAATPTGBALBLVLHLGSLRGVRVSSDA
TARPDGLDLAELPARGEWHTTALVRPSVEGEAGGLPFGPAEPGSGPARRLCAWRE
```

```
SAPRIITDIEDVEVHRTLRRSLDGLGAFAHLRPAAPRTHGRRRCPLVHGHPVRPRLAAV
LVHGTAAGPCTPR"
5183..5665
/note="SAV453"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68163.1"
/db_xref="GI:29604092"
/translation="MVTAVRARAGSAQACRARRSSOHSSEGVRRMDEKEFLRTV
AERTGLSRQEAADLTWATLETLARLSPGEARDLITELPEGLAEAVRGTGTDRIERFG
HHDSVRRVAERTMLKEEADRGVRTVLAVLREAISEKEFNDLMSQLGTFDSQALASAG
"
5810..6964
/note="SAV454"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68164.1"
/db_xref="GI:29604093"
/translation="MIMTADVPTSSGRSVRVALVADPGLPSEIALARALAPRLPDHLRR
AVGARVRWQVDTMTAPLSVSEQADVSDIAEAVRSHLDGEWDVAIFLDFPRLRLYP
ISEVDTALRAALISLPALGVRRRLRRVRQAVDVVRELVAHEETPRPEVICRPRE
PVDPTRPQPQPVVYVGLRGHTRLVAGMYANRPMRLFGSLSRALAGVFATATVLF
INSATWNLATAGAVQHAVITAGSAVALTWIIVDQHLWERTGLHPARHHPLYPNLV
TLVTIALVVVLAALPATLVAVSFLLLIDSLARKFIGRQVGVGYLFLAWFITAIAM
IGGAFGTSLRGDEKVSADAAVYRRQRQRMLRQMPRPERGERRGPDTTDG"
complement(6986..7498)
/note="SAV455"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68165.1"
/db_xref="GI:29604094"
/translation="MNPYAPASGIEQRPGSWLDEYWCSDIITADLAEGRLQRPVL
ATTARLAEAGRVLAVGPAQRVTWALGDGYMHQNAVFAGSPAFVLGSMASMTSALLG
NSARRAEANDAOAPRCGARTPGLHVDLVDVHGTMRLMSHMLGTSSTRRPSALRALG
ELVDLHVAL"
7548..8189
/note="SAV456"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68166.1"
/db_xref="GI:29604095"
/translation="MATLLRRVCPHEGVRIAAVAGFVLSRRPWTASVYRCVAVGSPQS
EDRREVGSTPVASASSRLRRCRRILSVRCPRRAVSQAVRSRFRGTTTFLGDLV
VYQRHDLARHVHVIDDVHPGLRGAAREVRIATHSLGGVIAVDMATANVPLMTES
LVTFGSQATFFHQVPVPGDLLCQMLAFMRKHQEAETGALPPETGFIAHAGQGW"
Alignment Scores:
Pred. No.: 2,91e-19 Length: 299175
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 1 Gaps: 0
US-09-917-376-5 (1-88) x AP005023 (1-299175)
QY 2 SerGlyGlyVallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 105341 TCGGGCGGCTCAAGTCTCTACAGAACCAACACTCTCGGCCACCACGACGCCATC 105400
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 105401 CGGCCAGGCTTCGATCGTCAACACCGGACGGCTCCCTCGACCTGTCCAAGGTACG 105460
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
Db 105461 GCCGCTACTACTTCACTCCCGGACAGCGGCTCGCCCAAGCGCTGGTGCAGACTAC 105520
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
```

[illegible]

```

CDS
620..3412
/gene="manA"
/codon_start=1
/transl_table=11
/product="multidomain beta-1,4-mannanase precursor"
/db_xref="GI:6651327"
/translation="WNRLIARLSGLMVLIAAMLAYVYKPEPAEAGHGMVFPATRT
YACYVDGKGVNGGDLNMINPACLDALAIISGNQVFNWFGNLIISNAGGSHREIIPDGK
LCGPTAFSGNMNQARTDMWTTLQPGATITVRNAWAPHGTYWLYVTRDWDPTQPL
KWSLEPTPFPSOVTPNPINSSGPDGAESYMOVLNKGGRHIIIMYWRSDSPEAFYN
CSDVYFGSPPIAYEGRDREGGTMITPPSGTPTPTPTPTPTPTPTPTPTPTPTPTPT
TPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
STQAGYFCDMAQICCSNIRAQFVLSQVSGADSYIELFTGSGIPAGNCTGEIQNR
IHTQMYNINTEDDSYNAQTWTGSPSTRITLYRNGVLVWGTEPGGSSPPTPTPTPT
PPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
GANAVRIIVLSNCRWTKIPASEVADIISOARTLGVRVAVLEVDHDTTGVEDAACSMT
TAVNTWIELKNVLAQENFVIYVNGNEPYGNNNYQNWVTDTRNAVALRAGINNTIM
VDAPNGQDMSFTMDNPATIFNADPQRLNLFVSIHMVGYDTAAEVQSYIESFVNRGL
PLVIGFQGHMSDGDNEQAIQVYAKQYNIIGFMSWNGSGGGVEYLDVMTFNANSP
TAWGTWFRNTAIGTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
KPHFRIVNRGTSSVPLSELTIRYWTVDGDKQVENCDAQVGCNLRGSEFVKLSGR
TCADYVYIETFTSGAGSLAPGASSGDIQVRINKNDWNTYNEANDYSYDPTKTSFADNW
RVLYRNGQLWGVPEP"
620..718
/gene="manA"
/evidence=not_experimental
719..3409
/product="multidomain beta-1,4-mannanase"
/gene="manA"
1349..1480
/note="Region: proline-threonine linker"
/evidence=not_experimental
1481..1927
/gene="manA"
/note="Region: cellulose-binding domain type IIib"
/evidence=not_experimental
1928..2032
/gene="manA"
/note="Region: proline-threonine linker"
/evidence=not_experimental
2033..2884
/gene="manA"
/note="Region: beta-1,4-mannanase catalytic domain"
/evidence=experimental
2885..2959
/gene="manA"
/note="Region: proline-threonine linker"
/evidence=not_experimental
2960..3409
/gene="manA"
/note="Region: cellulose-binding domain type IIib"
/evidence=not_experimental
3894..4400
/note="orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF22275.1"
/db_xref="GI:6651328"
/translation="WGASGVFSRTAONGYGLKARCKRRCVGNLGLHRLFVVDDDEF
RPFAGARKELGFSRQNLGEAQNFEISCTVFQPDKRRIRERNIGOLFVRFQASI
GPARLEINAEQARARKQASDQPLKNVILHPIDQHPYGVGERMDPFHKFEKQNS
RSILTRY"

ALIGNMENT SCORES:
Pred. No.: 1,3e-17 Length: 4567
Score: 225.50 Matches: 41
Percent Similarity: 67.44% Conservative: 17
Best Local Similarity: 47.67% Mismatches: 27
Query Match: 48.08% Indels: 1

ORIGIN

```

```

DB: 1 Gaps: 1
US-09-917-376-5 (1-88) x AFI63837 (1-4567)
QY 3 GlyGlyVallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLys 22
Db 58 GGAGCGCTGCTGTCGAGTAGTATCGCGCGGACACGACGCGGGCGCAACACGCTGAAG 117
QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 118 CCGATTTCGGATTGTGAACCGCGGACGTCGAGCGTGCCTGCTCGAGCTGACGATT 177
QY 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62
Db 178 CGGTACTGTTACACGGTGGAC---GGGACAAAGCCGAGGTGTTCAACTGCGACTGGGG 234
QY 63 AlaMetCysGlyGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 235 CAGTGGGTGTTGCGAACGTGCGCGGACGTTCTGAAGCTTTTCGACGGGCGGACGGGG 294
QY 83 AlaAspThrTyrLeuGln 88
Db 295 GCGGACTACTATATCGAG 312

RESULT 8
AF200304 3237 bp DNA linear BCT 01-DEC-2000
LOCUS
DEFINITION
Caldibacillus cellulovorans putative type IIib cellulose-binding
domain gene, partial cds; and beta-1,4-xylanase XynA precursor
(XynA) gene, complete cds.
ACCESSION
AF200304
VERSION
AF200304.1 GI:7385018
KEYWORDS
SOURCE
ORGANISM
Caldibacillus cellulovorans
Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;
Caldibacillus.
REFERENCE
1 (bases 1 to 3237)
Sunna, A., Gibbs, M.D. and Bergquist, P.L.
A novel thermostable multidomain 1,4-beta-xylanase from
'Caldibacillus cellulovorans' and effect of its xylan-binding
domain on enzyme activity
Microbiology 146 (Pt 11), 2947-2955 (2000)
11065373
PUBMED
2 (bases 1 to 3237)
Sunna, A., Gibbs, M.D. and Bergquist, P.L.
Direct Submission
Submitted (29-OCT-1999) Biological Sciences, Macquarie University,
North Ryde, Sydney, New South Wales 2109, Australia
JOURNAL
FEATURES
Location/Qualifiers
1..3237
/organism="Caldibacillus cellulovorans"
/mol_type="genomic DNA"
/db_xref="taxon:74586"
<1..-279
/notes="orf1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative type IIib cellulose-binding domain"
/protein_id="AAP61648.1"
/db_xref="GI:7385019"
/translation="QVGCNLRGSFVKLSGTGRTGADYVYIETFTSGAGSLAPGASSGD
IQVRINKNDWNTYNEANDYSYDPTKTSFADNWRVLYRNGQLWGVPEP"
384..3149
/gene="xynA"
384..3149
/gene="xynA"
/note="multidomain protein"
/codon_start=1
/transl_table=11
/product="beta-1,4-xylanase XynA precursor"

```





```
/EC_number="3.2.1.91"
3456..3472
/gene="cel148C"

terminator

ORIGIN

Alignment Scores:
Pred. No.: 9,26e-16 Length: 3509
Score: 209.50 Matches: 40
Percent Similarity: 70.11% Conservative: 21
Best Local Similarity: 45.98% Mismatches: 25
Query Match: 44.67% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-5 (1-88) x PAR488933 (1-3509)

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 2879 ACGGAAACCTGGAAGTGCAGTATCGCAATGAGGTCGCTTCAGTATGCAAT 2938

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 2939 ACGCCGCAATTTAAATCTGAAATAACAGGTACACGCGCAATTGATCTGAGTAAGGTGAAG 2998

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 2999 GTTCGTTATTATTTACGAAGGAC---ACGCGAGCGGACATGAGCTTCTGGTGTGACTAT 3055

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 3056 GCACAGCTGGCAGTCCCAATGTGAGCGCAGCTTTGCTGTAAATCCGCGGAAGGC 3115

Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 3116 ACGGCAGATACGTATCTGGAG 3136

RESULT 10
BSP133614
LOCUS Bacillus sp. BP23 celB gene. 4161 bp DNA linear BCT 09-MAY-2001
ACCESSION AJ133614
VERSION AJ133614.1 GI:4490765
KEYWORDS celB gene; cellulase; glycosyl hydrolase.
SOURCE Bacillus sp. BP-23
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1
AUTHORS Pastor,F.I., Pujol,X., Blanco,A., Vidal,T., Torres,A.L. and Diaz,P.
TITLE Molecular cloning and characterization of a multidomain
endoglucanase from Paenibacillus sp BP-23: evaluation of its
performance in pulp refining
JOURNAL Appl. Microbiol. Biotechnol. 55 (1), 61-68 (2001)
MEDLINE 21129642
PubMed 11234960
REFERENCE 2 (bases 1 to 4161)
AUTHORS Diaz,P.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1999) Diaz P., Microbiology, University of
Barcelona, Av.Diagonal 645, Barcelona-08028, SPAIN
FEATURES
source
1..4161
/organism="Bacillus sp. BP-23"
/mol_type="genomic DNA"
/strain="BP-23"
/db_xref="taxon:89769"
398..3525
/gene="celB"
-35_signal 398..404
/gene="celB"
-10_signal 437..442
/gene="celB"
RBS 514..519
/gene="celB"
CDS 532..3525

/gene="celB"
/EC_number="3.2.1.4"
/standard_name="endoglucanase"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="CAB38941.1"
/db_xref="GI:4490766"
/db_xref="GOA:Q92411"
/db_xref="SPTREMBL:Q92411"
/translations="MKGSMRRVAIVTLTSLAGLGSFMSMTGLFQADAAAGDHNYAEA
LQKATFYERQSRGSLPANNRVEWRGDSGMQDADGVGDLTGGHYDAGDHVKGFGPMA
ASATMLAWSVVEYADGEQAGLEIKONIRWATDYFKAKHTKPNELWQGVKAGNTDH
AWMPAEVQMQRNPSFKIDACSCPSDLAAETAALAAASIVFADSDPVYSKLLQHAH
EYLNFDATYRGKTYCTIDTAAAFYNSWTGYEDELAWGAMLYLATNDNAYLSKLSAA
DRWSTGGSANWPLYTWQGWDSKHYGACIILARTISLNMPEAKFKIQTSTERNLDYWT
VCTNGSRGKTYPGGLAWLDOMGSLRVAANAFAISFYVSDMVDSPVKSRYSQNEATSOI
NYILGNPQSSVYVGYSQSPQPHRTAHSSMNNEDIIPANRHLIYGNVGGPNA
SDQYTDIDIGVSVENATDYNAGTGLAKLKNLLYQNHQPLANFPAPPEVKGDEYFVS
AAVRSSGSNTEIRALLNNRSGPARGDQLSKYFLDLSEVYAAAGTSDVQVTVSS
SEGATYQPVVDAAKRIYAITANFSNTKIYPGEGNRYKEVQFRIITGPGAMNPAND
PSYONLTGTNPVKSNIIPVYDAGVKVSGQEPGTPVAVPAAPAGVQAVAGSNAPLAW
SASAGAVSVTKBAEYVSGGPTTVAAGVNGLYTNTGLTNGKTYYYVYVAVNSAGESP
ASVQSGMFOAATVPGAVTSLGTAQNQNLNLSWTAASGASSYTVQRAVAGTGYTDVA
TGLAVLNYNDTTALNGTSYRVAANASGTLNVLNVTSPGPPVSTGTLEUYTRSG
GSGNSNNAVTPQFNKNTGTQAILDLSTVKIRYFTKDGTELSFWCDYAOVGSANVOG
MFVAVNPAGKTADTYVEISFTSGAGSLAGAETGVIQTRFSKNWSAFDQSNDSYDA
SKTAPAAWNNKVTAYQNTQVWGLEP"
532..633
/sig_peptide
/gene="celB"

ORIGIN

Alignment Scores:
Pred. No.: 4,83e-15 Length: 4161
Score: 204.50 Matches: 38
Percent Similarity: 67.82% Conservative: 21
Best Local Similarity: 43.68% Mismatches: 27
Query Match: 43.60% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-5 (1-88) x BSP133614 (1-4161)

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 3067 ACAGGTACGCTTGGAGTGAGTATCGAAGCGGGGTTCCAGCAATTCAGCAATGCAGTT 3126

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 3127 ACTCCGCAATTCATCTGAAGAATACCGCACACAGCGCATTTGATCTGAGTACGTTGAAA 3186

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 3187 ATTCTGTTATTAATTTACCAAGAC---GGCACGAGGAGGAGTCTCTCTCTGTTGATTAAT 3243

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 3244 GCGCAGGTGGCTCAGCGAAGTACAGGAATGTTTGTAGCAGTGAATCCGCGCAAGGT 3303

Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 3304 ACAGCGATACGTATGTGGAG 3324

RESULT 11
E35142
LOCUS Truncated cellulase composition. 2029 bp DNA linear PAT 18-JUN-2001
DEFINITION E35142
ACCESSION E35142
VERSION E35142.1 GI:13018967
KEYWORDS JP 1999221086-A/44.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 2029)
```

AUTHORS		Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,	
TITLE		Hyu,M. and Daian,P.W.	
JOURNAL		Truncated cellulase composition	
COMMENT		Patent: JP 199221086-A 44 17-AUG-1999; CLARIANT INTERNATIONAL LTD	
OS		Artificial Sequence	
PN		JP 199221086-A/44	
PD		17-AUG-1999	
PF		21-SEP-1998 JP 1998283606	
PR		21-SEP-1997 US 08/932571	
PI		PAIJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI	
GURAHAMU K FARINTON,			
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONITOTSU WILLIAM			
PC		C12N15/09,C11D3/386,C12N1/21,C12N9/42/(C12N1/21,C12R1:19), PC	
(C12N9/42,C12R1:19),C12N15/00			
CC			
FH		Key	
FT		Location/Qualifiers	
FT		1..2029	
/organism='Artificial Sequence'			
FEATURES		Location/Qualifiers	
source		1..2029	
/organism="synthetic construct"			
/mol_type="genomic DNA"			
/db_xref="taxon:32630"			
ORIGIN			
Alignment Scores:			
Pred. No.:		2 6e-14 Length: 2029	
Score:		195.50 Matches: 38	
Percent Similarity:		60.23% Conservative: 15	
Best Local Similarity:		43.18% Mismatches: 34	
Query Match:		41.68% Indels: 1	
DB:		6 Gaps: 1	
US-09-917-376-5 (1-88) x E35142 (1-2029)			
QY	1	ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln	20
DB	1	ATGGGAAGTGGTGTGAAGGTACTGTACAAAGCAATGAGACAAGTCGACACAGGTCTCT	60
QY	21	ILeIysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal	40
DB	61	ATAAGCCGTGGTTTAAGATAGTAGTAATGGAGCGACGACAGTGTGATCTTAGCAGGGTT	120
QY	41	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp	60
DB	121	AAATGAAGATACATGGTACACATGGTGAATGGTGACAAAGCCACAGATGCGGTA---TGTGAC	177
QY	61	TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr	80
DB	178	TGGGCACAGTAGGGCGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGAGTG	237
QY	81	ProThrAlaAspThrTyrLeuGln	88
DB	238	AGTGGAGCGGATTATTACCTGGAG	261
RESULT 12			
E35143			
LOCUS		E35143	
DEFINITION		Truncated cellulase composition.	
ACCESSION		E35143	
VERSION		E35143.1 GI:13018968	
KEYWORDS		JP 199221086-A/45.	
SOURCE		unidentified	
ORGANISM		unclassified.	
REFERENCE		1 (Bases 1 to 2029)	
AUTHORS		Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,	
TITLE		Hyu,M. and Daian,P.W.	
JOURNAL		Truncated cellulase composition	
COMMENT		Patent: JP 199221086-A 45 17-AUG-1999; CLARIANT INTERNATIONAL LTD	
OS		Unidentified	

PN	JP 1999221086-A/45
PD	17-AUG-1999
PF	21-SEP-1998 JP 1998283606
PR	19-SEP-1997 US 08/932571
PI	PALJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI GURAHAMU K FARINTON,
PT	MORELAND DAVID GIBBUS,HYU MORGAN,DAIAN PURATONIOTSU WILLIAM C12N15/09,C11D13/386,C12N1/21.C12N9/42/(C12N1/21,C12R1:19),PC(C12N9/42,C12R1:19),C12N15/00
CC	Location/Qualifiers
FH	Key
FT	source
FEATURES	source
ORIGIN	source
Alignment Scores:	
Preld. No.:	2 6e-14 Length: 2029
Score:	195.50 Matches: 38
Percent Similarity:	60.23% Conservative: 15
Best Local Similarity:	43.18% Mismatches: 34
Query Match:	41.68% Indels: 1
DB:	Gaps: 1
US-09-917-376-5 (1-88) x E35143 (1-2029)	
Qy	1 ValSerGlyGlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20 :::
Dd	1 ATGGGAAGTGCTGTGAAGTACTGTACAGAACAATGAGCAAGTCGACACAGTTCT 60 
Qy	21 ILelyPProGlyLeuGlnleuValAsnThrGlySerSerValAspLeuSerThrVal 40 
Dd	61 ATAAGGCCGTGGTTTAAGATAGTGAATGGAGGCAGCACGAGCGTGTGATCTTAGCAGGGTT 120 
Qy	41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrleuValTyrAsnCysAsp 60 :::
Dd	121 AGATAAGATACATGGTACACAGTGGATGGTGCACACCACAGAGTCGGTA---TGTGAC 177 
Qy	61 TrpAlaAlaMetGlyCysGlyAsnIleAraGAlaSerPheGlySerValAsnProAlaThr 80 
Dd	178 TGGGCACAGATAGGGCCAGCAATGTGCATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237 
Qy	81 ProThrAlaAspThrTyrleuGln 88 
Dd	238 AGTGGAGCGCATATTACCTGGAG 261 
RESULT 13	
AF0780385S	6005 bp DNA linear BCT 11-FEB-2001
LOCUS	Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial cds; and CelE gene, complete cds.
DEFINITION	
ACCESSION	AF078042
VERSION	AF078042.1 GI:12743878
KEYWORDS	5 of 5
SEGMENT	Caldicellulosiruptor sp. Tok7B.1
SOURCE	Caldicellulosiruptor sp. Tok7B.1
ORGANISM	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE	1 (bases 1 to 6005)
AUTHORS	Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.
TITLE	Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL	Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE	20171169
PUBMED	10706665
REFERENCE	2 (bases 1 to 6005)
AUTHORS	Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.

Williams, D.P. and Bergquist, P.L.  
Direct Submission  
Submitted (27-OCT-1999) Biological Sciences, Macquarie University,  
Sydney, NSW 2109, Australia  
Location/Qualifiers  
source  
1. .6005  
/organism="Caldicellulosiruptor sp. Tok7B.1"  
/mol\_type="genomic DNA"  
/isolate="Tok7B.1"  
/db\_xref="taxon:80339"  
c1. .506  
/note="ORF6; multidomain"  
/codon\_start=3  
/transl\_table=11  
/product="glycosyl hydrolase 6"  
/protein\_id="AAK06393.1"  
/db\_xref="GI:12743884"  
/translation="GQPTWNGTYTCNPNLHVKIVDYXGTDLGITASLANALYYSAAT  
KEYGVSDBAARNLAKELDRMNLVDDKGLSAPKRGDGYKRFQEVYIPAGWTGKM  
FNGDVKGSGVGFIDIRSKYQDPDWQKLVSAYNAGEAPEFRYHFWAQCDIAIANATY  
BILFNGQ"  
misc\_feature  
c1. .503  
/note="Region: family 48 glycosyl hydrolase domain"  
617. .5872  
/note="multidomain glycosyl hydrolase"  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="CelB"  
/protein\_id="AAK06394.1"  
/db\_xref="GI:12743885"  
/translation="MQEKAIKRVVSVITALLVLTLSCPPGIMPVKYAGGTYNYGEA  
LQKTFYFQMSGLKPSWVRNWRGSDGLDDKVDGLDTGGMHDAGDHFVFNFLPMS  
YASMLGWAVYEDAFKYSKOLEHLNQIEWANDYFVKCHPSKYVYVQVGDPTVDH  
NFWGPAEVMQKRPAYKCDLSPASSVAETAASLAVASVWIKERNSOKAASYLOHAK  
DLFEADTRSDAGTYATGTYTSGFIDDLGAAWVLYIATNDSYLTAKSELSEY  
ANGNTWTCQDDVRYGTLIMLAKITKRELYKGAVERNLDDHWDRIITYPGMAVLTG  
WGSRYATTAFLACVADWSCDNKTKYLNFAKSDQIDYALGSTGRSFVVGFTNY  
PQPHRRNAHSWANSMKIPYHRHILYALVGGPDSDDSDNDITDYQNEVACDYN  
AGIVGALKALYOLYGEDIDDFKAIETPTNDEIFVESKFGNSQGNPTVEIVSYNRT  
GWPEVTKLSFYKVIDETELIQAQSPDVVDVYVLEGKISQPVYWDKKNLYVY  
LVDFSGTKIYPGGEVHEHKKQAQKISVPGQYFWDPTNDSYKGLSLEKKNKIIAYD  
NNNLWGLEPAGATSTPATSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT  
YNNNTSASTGSRIRFWKIVNGSGSSVDLSRVKIRYKVTVDGKQPSAVDWAQKIGAS  
NVTNFNKLSSGADYVLEVGSSGAGLOPKGTGDIQVRFNKNDNWNQADDDW  
SWLQSMTNVGENAKVTLVYDGLVWGQEPGATPATSTPATPTPTPTPTPTPTPTPTPT  
TPTVTSATPTPATSPVGSYWTPESEYKALKVWANGNLSPTVNLNPKKIENVTG  
TAVDLRVKRYTYTIDGATQSVASSINPAYIDVFKLGANAGGADYVIEIGFK  
SAGVLAAGQSTKEIRLSIQKSGSYNDSYRSANSYIENEKVTGYIDDLVWQGR  
EPGRNAQIKWYANGNLGSMTNLNPKIENVTGTAVDLSRVKRYTYTIDGATQSV  
VSVTSSINPAYIDVFKLGANAGGADYVIEIGFKSAGVLAAGQSTKEIRLSIQKSG  
GSYNDSYVRSATGYIENEKVTGYIDGAIWGREPGRGTPAGGVTPTPATPTST  
PT  
PARRUGNRLLTNNWNNNSAGSDWTHSSDDIMCYITMGITGNDKNNPAVSVKFEHQ  
SIKQNAYSAILTLQWYVAKGNGTQVSESTAPSWAEVFKPKDGLSLQDPVNDNY  
VYMDFINLYNKYSSSATGKIGYILNEDPLMFTTHPIHPQKVTCSSELINKSVE  
LAKVITLDPDAEIPGPAQVGVYTLQDAPQWQNGNHRWFLSWYLEOMKASDS  
FGKRLDLVDLHWYPEAQVGVYICFDGENSTRDVAIRMOAQRTPLDWPTVTKTQKG  
QITAGENSWINQWFEYLLPNIKADIDKYPGKLAITEFDYCGKDHISGIALAD  
VLGIFEGKGVYMAAGWSGSYAQAAYNIYLNIDKGRYSGTCSVAESTTDVENMPVY  
ASIEGDSFTYHILINRYDRKLKAEIKMNNTRYTGTGEIYGFDSSTQIRRMGVLS  
NIQNTTIEVENLTVYHIVLTSSK"  
725. .2053  
/note="Region: family 9 glycosyl hydrolase domain"  
2054. .2345  
/note="type IIc cellulose binding site"  
2627. .3106  
/note="type IIb cellulose binding site"  
3224. .4195  
/note="type IIb cellulose binding site"  
4298. .5869  
/note="Region: family 44 glycosyl hydrolase domain"  
2537. .2626

/note="PT-linker"  
3107. .3223  
/note="PT-linker"  
4196. .4295  
/note="PT-linker"  
ORIGIN  
Alignment Scores  
Pred. No.: 1.36e-13 Length: 6005  
Score: 194.50 Matches: 38  
Percent Similarity: 61.18% Conservative: 14  
Best Local Similarity: 44.71% Mismatches: 32  
Query Match: 41.47% Indels: 1  
DB: 1 Gaps: 1  
US-09-917-376-5 (1-88) x AF078038S5 (1-6005)  
QY 4 GlyVallyValGlnTyrIleValAsnAspSerAlaProGlyAspAsnGlnIleIysPro 23  
Db 2648 GGTGTGAAGGTACTGTACAAGAACAAATGAGACAAGTGGACACACAGGTTCTATAGGCGG 2707  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
Db 2708 TGGTTTAAGATAGTGAATGAGGAGCAGCAGCAGTGTGATCTTAGCAGGTTAAGATAAG 2767  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 2768 TACTGGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2824  
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 2825 ATAGGGGCAAGCAATGTGACATTTTGTGAAGCTTAGCAGCGGAGTGGTGGCGG 2884  
QY 84 AspThrTyrLeuGln 88  
Db 2885 GATTATACCTGGAG 2899  
RESULT 14  
E35100  
LOCUS Truncated cellulase composition.  
DEFINITION E35100  
ACCESSION E35100  
VERSION E35100.1 GI:13018925  
KEYWORDS JP 1999221086-A/2.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 6416)  
AUTHORS Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G.,  
Hyu, M. and Daiian, P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 1999221086-A 2 17-AUG-1999;  
CLARIANT INTERNATIONAL LTD  
COMMENT OS Unknown  
PN JP 1999221086-A/2  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 1998283606  
PR 19-SEP-1997 US 08/932571  
PI PAJJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI  
GURAHAMU K FARINTON,  
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONOTISU WILLIAM  
PC C12N15/09, C12D3/386, C12N1/21, C12N9/42// (C12N1/21, C12R1/19), PC  
(C12N9/42, C12R1/19), C12N15/00  
CC Key Location/Qualifiers  
FH source 1. .6416  
FT /organism="Unknown".  
FEATURES  
source  
1. .6416  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 03:07:10 ; Search time 122.068 Seconds  
(without alignments)  
3062.577 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKVQYKNNDSAPGDNQ.....IRASFGSVNPATPTADTVLQ 88

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame; p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114534\_28231/app\_query.fasta\_1.1429  
-DB=N Geneseq\_29Jan04 -QPMF=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CGN 1 1 758 @runat\_11052004\_114534\_28231 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04: \*  
1: geneseqn1980s: \*  
2: geneseqn1990s: \*  
3: geneseqn2000s: \*  
4: geneseqn2001as: \*  
5: geneseqn2001bs: \*  
6: geneseqn2002s: \*  
7: geneseqn2003as: \*  
8: geneseqn2003bs: \*  
9: geneseqn2003cs: \*  
10: geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	2289	7 ABZ77633	Abz77633 Nucleotid
2	466	99.4	3687	7 ABZ77634	Abz77634 Nucleotid
3	462	98.5	2869	7 ABZ77632	Abz77632 Nucleotid
4	462	98.5	2869	9 ADD22922	Add22922 Acidothe
5	436	93.0	3365	7 ABZ76162	Abz76162 A. cellu
6	231.5	49.4	2600	2 AAQ15178	Aaq15178 Portion o
7	195.5	41.7	2029	2 AAQ15178	Aaq15178 DNA seque
8	195.5	41.7	2029	6 AAD26568	Aad26568 Active ce

9	194.5	41.5	6415	2	AAZ55662	AAZ55662 DNA seque
10	194.5	41.5	6416	6	AAD26526	Active ce
11	192.5	41.0	11706	6	AAZ55661	DNA seque
12	192.5	41.0	11707	6	AAD26525	Active ce
13	188	40.1	1434	6	AAZ41028	CMCase ge
14	188	40.1	1488	6	AAZ41025	CMCase ge
15	188	40.1	2510	6	ABK53202	Bacillus
16	186	39.7	2175	2	AAQ49820	NK-1 cell
17	182.5	38.9	2977	2	AAQ13001	Endo1 gen
18	165	35.2	1438	3	AAZ45336	DNA encod
19	165	35.2	1482	2	AAZ90978	DNA encod
20	165	35.2	1482	2	AAZ31562	Pectate 1
21	165	35.2	5562	2	AAZ86625	C. thermo
22	157	33.5	1314	6	ABK73393	Bacillus
23	155.5	33.2	4059	5	AAH75059	Nucleotid
24	150.5	32.1	4056	5	AAH75060	Nucleotid
25	130.5	27.8	4452	7	ACA27429	Prokaryot
26	126.5	27.0	1303	9	ADC27473	Fusion pr
27	126.5	27.0	1747	9	ADC27474	Fusion pr
28	126.5	27.0	3489	9	ADC27475	Fusion pr
29	120.5	25.7	1775	2	AAQ13003	Endo3 gen
30	114	24.3	486	2	AAQ72917	Cellulose
31	114	24.3	486	2	AAV74072	C. cellul
32	114	24.3	499	2	AAZ24930	Clostridi
33	114	24.3	507	5	AAZ11042	Clostridi
34	114	24.3	768	4	AAF86248	DNA seque
35	114	24.3	768	6	ABK52403	E. coli c
36	114	24.3	984	5	AAZ11046	Chimeric
37	114	24.3	1030	5	AAZ11044	Clostridi
38	114	24.3	1288	5	AAZ11045	Clostridi
39	113	24.1	573	5	AAZ11043	Clostridi
40	113	24.1	1146	2	AAZ27550	Gaussia 1
41	113	24.1	1146	6	AAZ22201	Gaussia s
42	78.5	16.7	764	3	AAZ12881	Aspergill
43	78.5	16.7	3147	7	ACA25846	Prokaryot
44	77	16.4	8107	3	AAZ26216	Genomic D
45	76	16.2	96597	8	ADA02720	Human NFA

#### ALIGNMENTS

RESULT 1

ABZ77633

ID ABZ77633 standard; DNA; 2289 BP.

XX ABZ77633;

AC ABZ77633;

XX 03-JUN-2003 (first entry)

XX DE Nucleotide sequence of the ManA polypeptide.

XX KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
XX KW food; feed; paper pulp; biofuel; mannanase; gene; ss.

XX OS Acidothermus cellulolyticus.

XX PH Key Location/Qualifiers

XX FT CDS 1..2289

XX FT /\*tag= a

XX FT /product= "ManA"

XX PN WO2003012110-A1.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023819.

XX PR 28-JUL-2001; 2001WO-US023819.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

```

DR WPI: 2003-248182/24.
XX P-PSDB; ABP73022.
PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
XX Example 1; Page 23; 46pp; English.
XX
XX The present sequence encodes ManaA, a thermostable mannanase A polypeptide
XX derived from Acidothermus cellulolyticus. ManaA is a member of the
XX glycoside hydrolase family of enzymes. ManaA is useful for reducing
XX hemicellulose in a starting material to simpler carbohydrate units, and
XX ultimately to sugars which are useful in the food, feed, paper pulp, and
XX biofuels industries. It is useful for the processing of food and in food
XX stuffs as bulking agents, and for the degradation of mannanase. ManaA is
XX also useful to raise polyclonal and monoclonal antibodies that are useful
XX in purifying ManaA, or detecting ManaA polypeptide expression, and as well
XX as reagent tools for characterizing the molecular actions of ManaA
XX polypeptides
XX
XX Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 4.13e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
XX
US-09-917-376-5 (1-88) x ABZ77633 (1-2289)
QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGTGAGTCAAGAACAAATGATTCGGCGCGGTGATAACCGAG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTGTGTCGTGGTGGATTGTGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGCGGTACTGTGTTTCAACCGGGATGGTGGTGTGTCGACACTGTGTTACAACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyValAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGGTGTGGGAATATCCGGCTCTGTTCCGGTGAACCCGCGGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626
XX
RESULT 2
ABZ77634
ID ABZ77634 standard; DNA; 3687 BP.
XX
XX ABZ77634;
AC ABZ77634;
XX
XX 03-JUN-2003 (first entry)
XX
XX Nucleotide sequence of the GuxA polypeptide.
XX
XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile;
XX gene; ss.
XX
XX Acidothermus cellulolyticus.
XX
XX Key Location/Qualifiers
XX CDS 1..3687
XX /*tag= a
XX /product= "GuxA"
XX

```

---

```

XX WO2003012109-A1.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001WO-US023817.
XX
XX 28-JUL-2001; 2001WO-US023817.
XX
XX (WIDE ) MIDWEST RES INST.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
XX WPI; 2003-239526/23.
XX
XX P-PSDB; ABP73029.
XX
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX cellulolyticus, useful for reducing cellulose in a starting material, and
XX for the conversion of biomass to biofuels and biofuel additives.
XX
XX Example 1; Page 23-24; 47pp; English.
XX
XX The present sequence encodes a GuxA polypeptide. GuxA is thermostable
XX cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX GuxA is useful for reducing cellulose in a starting material such as
XX agricultural biomass to sugars. This is useful in biofuel production.
XX GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX additives, in detergents, pulp and paper processing, food and feed
XX processing, and in textile process. GuxA is also useful for raising
XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX or detecting GuxA polypeptide expression, as well as reagent tools for
XX characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1.79e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0
DB: 7 Gaps: 0
XX
US-09-917-376-5 (1-88) x ABZ77634 (1-3687)
QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGTGAGTCAAGAACAAATGATTCGGCGCGGTGATAACCGAG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACGGGGTGTGTCGTGGTGGATTGTGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGCGGTACTGTGTTTCAACCGGGATGGTGGTGTGTCGACACTGTGTTACAACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyValAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGGTGTGGGAATATCCGGCTCTGTTCCGGTGAACCCGCGGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013
XX
XX RESULT 3
XX ABZ77632
ID ABZ77632 standard; DNA; 2869 BP.
XX
XX ABZ77632;
AC ABZ77632;
XX
XX 03-JUN-2003 (first entry)
XX

```

```

DE Nucleotide sequence of the avicelase AvIII.
KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose; gene; ss.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2869
FT /tag= a
FT /partial
FT /product= "AvIII"
FT /transl_except= (pos:2869,aa:Xaa)
FT /note= "Xaa is an unspecified residue"
XX
PN WO2003012090-A2.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023818.
XX
PR 28-JUL-2001; 2001WO-US023818.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI: 2003-248177/24.
DR P-PSDB; ABP73015.
XX
PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-AvIII antibodies for
PT purifying recombinant AvIII polypeptides from genetically engineered
PT host cells.
XX
PS Claim 3; Page 24; 44pp; English.
XX
CC The present sequence encodes a thermostable avicelase polypeptide,
CC designated AvIII. AvIII is a member of the glycoside hydrolase family
CC of enzymes, and is a cellulase. AvIII is useful in the conversion of
CC biomass to biofuels and biofuel additives. It may be useful in the
CC production of detergents, pulp and paper processing, food and feed
CC processing and in textile processes. The thermostable AvIII peptide is
CC useful in the degradation of cellulose, and in generating specific anti-
CC AvIII antibodies that are useful in purifying recombinant AvIII
CC polypeptides from genetically engineered host cells, in detecting AvIII
CC polypeptide expression, as well as a reagent tool for characterizing the
CC molecular actions of the polypeptide. The AvIII polynucleotide is useful
CC as a source of probes or primers in various diagnostic assays
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 4.09e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 7 Gaps: 0

US-09-917-376-5 (1-88) x AB277632 (1-2869)
QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 2605 GTGTCGGGTGGGTGAAGGTGCAGTATAAGATAATGATTCGGCGCGGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 2665 ATCAAGCCGGGTGGTGCAGGTGGTGAATACCGGGTCGTTCGGTGGATTTGTCGACGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

```

---

```

Db 2725 ACGTGGGTACTGGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTAACACTGTGAC 2784
QY 61 TTPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 2785 TGGGGCGCGATCGGGTGTGGGAATATCCGCGCCTCGTTCGGTTCGGTGAACCCCGCGACG 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 4
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase AvIII DNA.
XX
KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;
KW glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2868
FT /tag= a
FT /product= "AvIII"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI: 2003-810853/76.
DR P-PSDB; ADD22921.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
PS Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase AvIII.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 4.09e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0

```

DB: 9 Gaps: 0

US-09-917-376-5 (1-88) x ADD22922 (1-2869)

QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspGln 20  
 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGATAATGATTCGGCGCGGTGATATCAG 2664

Db

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 2665 ATCAAGCCGGTTTCAGGTGCTGAATACCGGGTCGTCGGTGGATTGTCGACGGTG 2724

Db

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 2725 ACGGTGCGGTACTGGTTTCACCGGGATGTTGGTCTGTCGACACTGTTGACACTGTGAC 2784

Db

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 2785 TGGGGCGGCGATCGGGTGTGGGAATATCCGGCTCGTTCGGTTCGGTGAACCGCGGACG 2844

Db

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 2845 CCGAGCGCGGACACTCTGTCGAG 2868

Db

RESULT 5  
 ID ABZ76162 standard; DNA; 3365 BP.

AC ABZ76162;

XX

DT 29-MAY-2003 (first entry)

XX

DE A. cellulolyticus Gux1 protein encoding DNA.

XX

KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 biofuel; detergent; pulp; paper processing; feed processing; textile;  
 cellulase; gene; ds.

KW

XX

OS Acidothermus cellulolyticus.

XX

XX

PH Key Location/Qualifiers

FT CDS 1..3365

FT /tag= a

FT /product= "Gux1"

FT /transl\_except= (pos: 682..683, aa: Pro)

FT /note= "this codon has an apparent one nucleotide  
 basepair deletion which alters the reading frame"

XX

PN WO2003012095-A1.

XX

XX

PD 13-FEB-2003.

XX

PF 28-JUL-2001; 2001WO-US023820.

XX

PR 28-JUL-2001; 2001WO-US023820.

XX

PA (WIDE ) MIDWEST RES INST.

XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 Lantz Mccarter S;

PI

XX

DR WPI; 2003-300494/29.

DR

XX

P-PSDB; ABP71656.

XX

XX

PT New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 useful in the degradation of cellulose to biofuels.

PT

PS Disclosure; Page 22-23; 44pp; English.

XX

XX

CC The invention relates to a thermal tolerant Gux1 peptide from A.  
 cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 binding domain type III, and a carbohydrate binding domain type II. The

CC

CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 for conversion of biomass to biofuel additives. It is used in detergents,  
 pulp and paper processing, food and feed processing, and in textile  
 processing. It can also be used alone or in combination with other  
 cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
 CC DNA

XX

SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.56e-44 Length: 3365

Score: 436.00 Matches: 81

Percent Similarity: 97.62% Conservative: 1

Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.96% Indels: 0

DB: 7 Gaps: 0

US-09-917-376-5 (1-88) x ABZ76162 (1-3365)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 112 CTCAAAGCCGAGTATAGAAACAATGATTCGGCGCGAGTGACACCAAGATCAACCGGT 171

Db

QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
 172 CTCAGTTGGTGAATACCGGGTCGTCTCGTGGTGGATTTCGACGGTGACGGTGGCGTAC 231

Db

QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
 232 TGGTTCACCGGGATGTCGGTCTGTCACACTGTCGTGTACAACTGTGCTGGCGCGATG 291

Db

QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
 292 GGGTGTGGGAATATCCGGCTCGTTCGGTTCGGTGAACCGCGGACCGCGCGGAC 351

Db

QY 85 ThrTyrLeuGln 88

Db

352 ACCTACCTGCAG 363

RESULT 6  
 AAQ15178

ID AAQ15178 standard; DNA; 2600 BP.

XX

AC AAQ15178;

XX

DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 17-MAR-1992 (first entry)

XX

DE Portion of pAEC-1 contg. cellulase AE-1 gene.

XX

KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.

XX

OS Aeromonas sp.

XX

PH Key Location/Qualifiers

FT CDS 192..2540

FT /tag= a

FT /product= "cellulase AE-1"

XX

PN JP03251174-A.

XX

PD 08-NOV-1991.

XX

PF 28-FEB-1990; 90JP-00045465.

XX

PR 28-FEB-1990; 90JP-00045465.

XX

PA (OJIP ) OJI PAPER CO.

XX



DR WPI: 1991-373412/51.  
 DR P-PSDB; AAR15625.  
 XX  
 PT Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
 PT pH when carboxymethylcellulose is used as substrate.  
 XX  
 PS Claim 3; Fig 2; 8pp; Japanese.  
 XX  
 CC The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene  
 CC (obtd. from Aeromonas) into pUC18. The gene can be used to produce  
 CC recombinant enzyme which is used for the effective utilis- ation of  
 CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and  
 CC also for the detergent and denking of waste paper. (Updated on 25-MAR-  
 CC 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.62e-18 Length: 2600  
 Score: 231.50 Matches: 42  
 Percent Similarity: 66.67% Conservative: 16  
 Best Local Similarity: 48.28% Mismatches: 28  
 Query Match: 49.36% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAQ15178 (1-2600)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 21  
 Db 2082 AGCGGGGATTTGGCGGTTCAGTACAAAACCGGAGATGCAATGCGCGCAGGTTTC 2141

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
 Db 2142 AAGCTCTATTTCATATCGTGAACAGGGGCTGCGCAGTGCCTTTAAGCGAGCTGAGT 2201

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 2202 CTGCGATATTATTTTCACAGCTGAC--GGCAATGACCAACTGCAATACAAATTTGACTGG 2258

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
 Db 2259 GCATGTGTAGGTGCTCCAACTGAACGGGGCTTTCTGTAATAATGAATCCGGCAAAGCG 2318

QY 82 ThrAlaAspThrTyrLeuGln 88  
 Db 2319 AACGCCGATACCTACTTGGAG 2339

RESULT 7  
 ID AAX55660  
 AC AAX55660;  
 XX  
 XX AAX55660 standard; DNA; 2029 BP.  
 DT 30-JUL-1999 (first entry)  
 DE  
 DE DNA sequence encoding truncated cellulase Cel E3/B5.  
 XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN EP921188-A2.  
 XX  
 XX 09-JUN-1999.  
 PD  
 PD 15-SEP-1998; 98EP-00810919.  
 PF  
 PF 19-SEP-1997; 97US-00932571.  
 PR  
 PR (CLRN ) CLARIANT FINANCE BVI LTD.  
 PA  
 XX

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 DR WPI: 1999-315403/27.  
 DR P-PSDB; AAY13494.  
 XX  
 PT New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 XX  
 PS Disclosure; Page 41-42; 65pp; English.  
 XX  
 CC The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkalophilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3.53e-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservative: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.68% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55660 (1-2029)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1 ATGGGAAGTGGTGTGAAGGTACTGTACAGAACATGACAGACAGTGCAGCAGGTTCT 60

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 61 ATAAGGCCGCTGGTTAAGATAGTGAATGGAGCGCAGCAGTGTTCATCTTAGCAGGGTT 120

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 121 AAGATAAGATACCTGGTACACAGTGGATGGTGGACAGCCACAGAGTGGCGTA---TGTGAC 177

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 178 TGGGCACAGATAGGGCGCAAGCAATGTGCACATTCATTTTGTGAAGCTTAGCAGCGGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 238 AGTGGAGCGGATTATTACCTGGAG 261

RESULT 8  
 ID AAD26568  
 AC AAD26568;  
 XX  
 XX AAD26568 standard; DNA; 2029 BP.  
 DT 26-MAR-2002 (first entry)  
 DE  
 DE Active cellulase hybrid protein, E3/B5 DNA.  
 XX  
 XX Active cellulase protein; alkalophilic; textile processing; proteinase;

KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.  
 XX Unidentified.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT 1..11851  
 CDS  
 FT /\*tag= a  
 FT /product= "E3/B5 hybrid protein"  
 XX  
 XX US62943366-B1.  
 XX  
 XX PD 25-SEP-2001.  
 XX  
 XX PF 19-AUG-1998; 98US-00136574.  
 XX  
 XX PR 19-SEP-1997; 97US-00932571.  
 XX  
 XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 XX DR WPI; 2002-081780/11.  
 DR P-PSDB; AAE16325.  
 XX  
 XX New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.  
 XX  
 XX PS Disclosure; Col 71-74; 61pp; English.  
 XX  
 XX The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the  
 CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is E3/B5 hybrid protein DNA  
 XX  
 XX SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3,536-14 Length: 2029  
 Score: 195.50 Matches: 38  
 Percent Similarity: 60.23% Conservative: 15  
 Best Local Similarity: 43.18% Mismatches: 34  
 Query Match: 41.68% Indels: 1  
 DB: 6 Gaps: 1  
 US-09-917-376-5 (1-88) x AAD26568 (1-2029)  
 QY 1 ValSerGlyValIysValGlnTyrIysAsnAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1 ATGGGAAGTGGTGTGAAGGTACTGTACAAAGCAATGAGACCAAGTCGCGACAGGTTCT 60  
 QY 21 IleLysProGlyLeuGlnLeuValThrGlySerSerValAspLeuSerThrVal 40  
 Db 61 ATAAGCCCGTGTAAAGATAGTGAATGAGGACGACGAGTGTGATCTTAGCAGGGTT 120  
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 121 AGATAAGATATCTGTACACAGTGGATGGTGTGACAGCCACAGATCGCGTA---TGTGAC 177  
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGAGTG 237

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 238 AGTGGAGCGGATTATTACTGGAG 261  
 RESULT 9  
 AAX55662  
 ID AAX55662 standard; DNA; 6415 BP.  
 XX  
 AC AAX55662;  
 XX  
 DT 30-JUL-1999 (first entry)  
 XX  
 DE DNA sequence encoding truncated cellulases.  
 XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN BP921188-A2.  
 XX  
 XX PD 09-JUN-1999.  
 XX  
 XX PF 15-SEP-1998; 98EP-00810919.  
 XX  
 XX PR 19-SEP-1997; 97US-00932571.  
 XX  
 XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
 XX  
 XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
 PI Morgan H, Williams DP;  
 XX  
 XX DR WPI; 1999-315403/27.  
 DR P-PSDB; AAY13493.  
 XX  
 XX New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 PT  
 PT Disclosure; Page 24-25; 65pp; English.  
 XX  
 XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkalophilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 XX SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;  
 Alignment Scores:  
 Pred. No.: 2,056-13 Length: 6415  
 Score: 194.50 Matches: 38  
 Percent Similarity: 61.18% Conservative: 14  
 Best Local Similarity: 44.71% Mismatches: 32  
 Query Match: 41.47% Indels: 1  
 DB: 2 Gaps: 1  
 US-09-917-376-5 (1-88) x AAX55662 (1-6415)



XX New truncated cellulase proteins, useful in detergents and for producing  
 PT 'stonewashed' denim.  
 XX  
 XX  
 PS Disclosure; Page 20-23; 65pp; English.  
 XX  
 XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkaliphilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions

XX  
 SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 7,82e-13 Length: 11706  
 Score: 192.50 Matches: 37  
 Percent Similarity: 62.35% Conservative: 16  
 Best Local Similarity: 43.53% Mismatches: 31  
 Query Match: 41.04% Indels: 1  
 DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55661 (1-11706)

QY 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
 DB 4038 GGTTTGAAGTACTATACAAAGCAATGAGCAAGTGGCAGCAAGTCTTATAAGCGC 4097  
 QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
 DB 4098 TGGTTTAAGATAGTGAATGAGGAGCAGCAGTGTGATCTTAGCAGGTTAAGATAAGA 4157  
 QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
 DB 4158 TACTGGTACACAGTGGATGGTGACAGCCACAGAGTGGCGGTA---TGTGACTGGGCACAG 4214  
 QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
 DB 4215 ATAGGGGCAAGCAATGTGACATTCATTTGTGAAGCTGAGCAGCGAGTGGAGCG 4274  
 QY 84 AspThrTyrLeuGln 88  
 DB 4275 GATTATTACTTGGAG 4289

RESULT 12

AAD26525

ID AAD26525 standard; DNA; 11707 BP.

XX

AC AAD26525;

XX

DT 26-MAR-2002 (first entry)

DE

XX Active cellulase protein, celB gene.

KW

XX Active cellulase protein; alkaliphilic; textile processing; proteinase;

KW

XX detergent additive; stonewashed appearance; cotton-containing denim;

KW

XX CelB5; thermophilic; commercial detergent; celB gene; ds.

XX

OS Unidentified.

XX

PH

Key Location/Qualifiers

CDS 6255..10535

FT /\*tag= a

FT /product= "CelB protein"

FT misc\_feature 8601..10532

FT /\*tag= b

FT /product= "DNA encoding B4/5 protein"

FT misc\_feature 9255..10526

FT /\*tag= c

FT /product= "DNA encoding B/5 protein"

XX

PN US6294366-B1.

XX

PD 25-SEP-2001.

XX

PF 19-AUG-1998; 98US-00136574.

XX

PR 19-SEP-1997; 97US-00932571.

XX

PA (CLRN ) CLARIANT FINANCE BVI LTD.

XX

PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;

PI Morgan H, Williams DP;

XX

DR WPI; 2002-081780/11.

DR P-PSDB; AAE16323.

XX

PT New cellulase active protein, useful in textile processing or commercial

PT

PT detergents, e.g. for improving the feel or appearance of cotton-

PT

PT containing fabrics, is stable under conditions of alkaline pH and

XX

XX elevated temperatures.

XX

PS Disclosure; Col 27-38; 61pp; English.

XX

CC The present invention relates to a cellulase active protein, which is

CC

CC substantially free of proteinases of native thermophilic and

CC

CC alkaliphilic origin, where the cellulase active protein consists of the

CC

CC CelB5 amino acid sequence. The cellulase active protein is useful for

CC

CC treating cellulosic materials including cotton-containing fabrics, as

CC

CC detergent additives. The cellulase active protein is also useful for

CC

CC improving the feel and/or appearance of cotton-containing fabrics, for

CC

CC removing surface fibers from cotton-containing knits or for imparting

CC

CC stonewashed appearance to cotton-containing denims. The present proteins

CC

CC are stable under condition of alkaline pH and elevated temperatures, thus

CC

CC suitable for textile processing and in commercial detergents. The present

CC

CC sequence is celB gene

XX

SQ Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,83e-13 Length: 11707

Score: 192.50 Matches: 37

Percent Similarity: 62.35% Conservative: 16

Best Local Similarity: 43.53% Mismatches: 31

Query Match: 41.04% Indels: 1

DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x AAD26525 (1-11707)

QY 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

DB 4038 GGTTTGAAGTACTATACAAAGCAATGAGCAAGTGGCAGCAAGTCTTATAAGCGC 4097

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

DB 4098 TGGTTTAAGATAGTGAATGAGGAGCAGCAGTGTGATCTTAGCAGGTTAAGATAAGA 4157

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63

DB 4158 TACTGGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

Db	4215	ATAGGGCAAGCAATGTGACATTTTGTGAAGCTGAGCAGCGGAGTGAGTGAGCG	4274
Qy	84	AspThrTyrLeuGln 88	
Db	4275	GATTATTACTTGGAG	4289
RESULT 13			
AAAL41028			
ID	AAAL41028	standard; DNA; 1434 BP.	
XX			
AC	AAAL41028;		
XX			
DT	11-OCT-2002	(first entry)	
XX			
DE	CMCase	gene with additional His encoding sequences.	
XX			
KW	Protein	surface-displayed; genetic carrier; spore; monoclonal variant;	
KW	high-throughput	screening; antibody; vaccine; enzyme; gene; ds.	
XX			
OS	Unidentified.		
XX			
XX	Key	Location/Qualifiers	
PH		1..1434	
FT	CDS		
FT		/*tag= a	
FT		/product= "CMCase protein"	
XX			
PN	WO200255561-A1.		
XX			
PD	18-JUL-2002.		
XX			
PF	15-JAN-2002; 2002WO-KR000059.		
XX			
PR	15-JAN-2001; 2001KR-00002156.		
XX			
PA	(GENO-) GENOFOCUS CO LTD.		
XX			
PI	Pan JG, Choi SK, Jung HC;		
XX			
PI	WPI; 2002-590661/63.		
XX			
DR	P-PSDB; AAO22444.		
XX			
PT	Methods	for surface display of proteins, e.g. displaying a protein on the	
PT	surface of spores, improving	a protein or isolating a substance, by	
PT	transforming a host cell	harboring a genetic carrier with spore or virus	
PT	with the	vector library.	
XX			
PS	Example 5; Page 109-113; 118pp; English.		
XX			
CC	The invention	relates to methods for preparing a protein surface-	
CC	displayed on a genetic	carrier, improving a protein or isolating a	
CC	substance in a mixture	comprising transforming a host cell harbouring a	
CC	genetic carrier	consisting of a spore or virus with the vector library.	
CC	The method	is useful for surface display of proteins, particularly for	
CC	displaying a protein	on the surface of e.g. spores, for improving a	
CC	protein, or for isolating	a substance. These methods are useful in	
CC	obtaining monoclonal	variants from a large library, or in high-throughput	
CC	screening of antibodies	for use in therapy (e.g. as vaccines), diagnosis	
CC	or analysis. This	polynucleotide sequence represents the CMCase gene with	
CC	additional His	encoding sequences relating to the invention	
XX			
SQ	Sequence 1434 BP; 493 A; 293 C; 341 G; 307 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1.94e-13	Length:	1434
Score:	188.00	Matches:	34
Percent Similarity:	61.18%	Conservative:	18
Best Local Similarity:	40.00%	Mismatches:	31
Query Match:	40.09%	Indels:	2
DB:	6	Gaps:	1
US-09-917-376-5 (1-88) x	AAAL41028 (1-1434)		
Qy	4	GlyValIysValGIntyrlYlsAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23

```

XX SQ Sequence 1488 BP; 501 A; 297 C; 357 G; 333 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2,04e-13 Length: 1488
XX Score: 188.00 Matches: 34
XX Percent Similarity: 61.18% Conservative: 18
XX Best Local Similarity: 40.00% Mismatches: 31
XX Query Match: 40.09% Indels: 2
XX DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x AAL41025 (1-1488)
QY 4 GlyValLysValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 1048 GGTATTTCGTACATACAGACGAGGGGATGGAGTATGACACGACCAACCAATCCGTCG 1107
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 1108 CAGCTTCAAAATAAATAACGGCAATACCGTTGATTTAAAGATGTCACGTCGCGT 1167
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 1168 TACTGGTATACCGGAAACCAAGGCCCAAAAGCTT-----GACTGTGACTACGGCAG 1221
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 1222 CTTGGATGCGGCAATGTGACATACAAAGTTTGTGACGTTCGATTAACCAAGCAAGGTGCA 1281
QY 84 AspThrTyrLeuGln 88
DB 1282 GATACCTATCTGGAA 1296

RESULT 15
ID ABK53202 standard; DNA; 2510 BP.
XX AC ABK53202;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus subtilis spore coat protein cotG-CMCCase fusion gene.
KW Spore coat protein; spore surface; delivery system; bioconversion;
KW recombinant live vaccine; resistance property; protein microarray;
KW whole cell absorber; secretion signal; target signal; cell membrane;
KW surface display motif; cotG; CMCCase; carboxymethyl cellulase; gene; ds.
OS Bacillus subtilis.
OS Synthetic.
OS Chimeric.
XX FH Key Location/Qualifiers
XX FT Promoter 1...460
XX FT /tag= a
XX FT /note= "CotG promoter"
XX FT CDS 461..2494
XX FT /tag= b
XX FT /partial
XX FT /product= "Bacillus subtilis spore coat protein cotG-
XX FT carboxymethyl cellulase (CMCCase) fusion protein"
XX FT /note= "This sequence lacks a start codon"
XX FT misc_feature 461..1045
XX FT /tag= c
XX FT /note= "CotG structural gene"
XX FT misc_feature 1046..1084
XX FT /tag= d
XX FT /note= "Linker"
XX FT misc_feature 1085..2491
XX FT /tag= e
XX FT /note= "CMCCase structural gene"
XX PN W0200246388-A1.

```

```

XX PD 13-JUN-2002.
XX
XX 07-DEC-2001; 2001WO-KR002124.
XX
XX 08-DEC-2000; 2000KR-00074835.
XX
XX (GENO-) GENOFOCUS CO LTD.
XX
XX PI Pan J, Choi S, Jung H;
XX
XX WPI: 2002-463830/49.
XX P-PSDB; AAU98063.
XX
XX Displaying a protein of interest on spore surface, by transforming a host
XX cell with a vector comprising a gene encoding a spore coat protein and a
XX gene encoding a protein of interest, useful for preparing protein
XX microarray.
XX
XX Example 4; Page 145; 158pp; English.
XX
XX The present invention relates to a new method for displaying a protein of
XX interest on a spore surface. The method of the invention involves
XX preparing a vector comprising a gene construct containing a gene encoding
XX a spore coat protein and a gene encoding a protein of interest,
XX transforming a host cell with the vector, displaying the protein of
XX interest on a surface of a spore of the host cell, and recovering the
XX spore. The method is useful for producing a delivery system for
XX recombinant live vaccine. The method is also useful for improving a
XX protein of interest using a resistance property of spore, in
XX bioconversion reactions, for preparing protein microarray, for producing
XX an antibody to antigen in vertebrates, and for preparing a whole cell
XX absorber. The spore coat proteins used circumvent a necessity for passage
XX across cell membrane, so that they do not need secretion signal and
XX target signal which are prerequisites of surface display motif, therefore,
XX ensuring a surface display of protein, such as beta-galactosidase, in
XX orderly fashion, which is difficult to pass across cell membrane. The
XX present nucleic acid sequence encodes the Bacillus subtilis spore coat
XX protein cotG-carboxymethyl cellulase (CMCCase) fusion protein of the
XX invention
XX
XX SQ Sequence 2510 BP; 869 A; 528 C; 521 G; 592 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3.97e-13 Length: 2510
XX Score: 188.00 Matches: 34
XX Percent Similarity: 61.18% Conservative: 18
XX Best Local Similarity: 40.00% Mismatches: 31
XX Query Match: 40.09% Indels: 2
XX DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x ABK53202 (1-2510)
QY 4 GlyValLysValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 2051 GGTATTTCGTACATACAGACGAGGGGATGGAGTATGACACGACCAACCAATCCGTCG 2110
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 2111 CAGCTTCAAAATAAATAAATAACGGCAATACCGTTGATTTAAAGATGTCACGTCGCGT 2170
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 2171 TACTGGTATACCGGAAACCAAGGCCCAAAAGCTT-----GACTGTGACTACGGCAG 2224
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 2225 CTTGGATGCGGCAATGTGACATACAAAGTTTGTGACGTTCGATTAACCAAGCAAGGTGCA 2284
QY 84 AspThrTyrLeuGln 88
DB 2285 GATACCTATCTGGAA 2299

```

Search completed: May 12, 2004, 05:39:39  
Job time : 137.068 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:33:52 ; Search time 826.547 Seconds  
(without alignments)  
3179.336 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 459  
Sequence: 1 VSGGVKVOYKNDSAPGDNQ.....IRASFGSVNPATPTADTYLV 88

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgm2\_1/USPRO\_pool/US09917376/runat\_11052004\_114535\_28249/app\_query.fasta\_1.1429  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CGN 1 1 5334 @runat\_11052004\_114535\_28249 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81	17.3	584	14	CF480372	CF480372 POL1 65 H
2	79.5	17.0	754	12	BQ047524	BQ047524 EST556642
3	78.5	16.7	768	14	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	13	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	9	AA043716	AA043716 zk61f10.8
6	76	16.2	606	10	BE362891	BE362891 DGI 90 C1
7	76	16.2	641	14	CD226529	CD226529 CCL1 46 E
8	76	16.2	657	10	BE362820	BE362820 DGI 89 C0
9	76	16.2	676	14	CF487649	CF487649 POL1_45 C
10	76	16.2	691	12	BG739392	BG739392 EMI_82_D0
11	76	16.2	724	28	BZ616323	BZ616323 i958e08.9
12	76	16.2	814	12	B1218457	B1218457 602937951
13	76	16.2	905	28	BZ545807	BZ545807 OGALR72TM
14	76	16.2	934	29	CG270741	CG270741 OGWB33TH
15	76	16.2	1001	29	CG270748	CG270748 OGWB33TV
16	75.5	16.1	657	10	BF513128	BF513128 UI-H-BW1
17	75.5	16.1	1058	14	CK231562	CK231562 ILLUMIGEN
18	74.5	15.9	465	10	AW491464	AW491464 UI-M-BH3
19	74.5	15.9	613	14	CA595709	CA595709 wpalc.pk0
20	74.5	15.9	650	28	BZ894843	BZ894843 Hg4_0131
21	74	15.8	302	14	CF076729	CF076729 QHK12L22.
22	74	15.8	653	12	B1664037	B1664037 603289338
23	73.5	15.7	261	10	BF875908	BF875908 RC3-ET013
24	73.5	15.7	816	14	CB293980	CB293980 UCRCS01_0
25	73.5	15.7	819	14	CB293084	CB293084 UCRCS01_0
26	73.5	15.7	1039	12	BG334904	BG334904 607461219
27	73.5	15.7	1237	13	BQ065850	BQ065850 AGENCOURT
28	73	15.6	357	14	CF078232	CF078232 QHK1J23.Y
29	73	15.6	360	13	C64738	C64738 C64738 YuJ1
30	73	15.6	593	14	CF622108	CF622108 laf13d06.
31	73	15.6	776	29	BK209426	BK209426 Dario rer
32	73	15.6	861	28	BH319803	BH319803 CH230-115
33	73	15.6	894	29	CG077853	CG077853 PUFJKJ97B
34	73	15.6	924	29	CG077856	CG077856 PUFJKJ97D
35	72.5	15.5	380	12	BG816375	BG816375 dad35b08.
36	72.5	15.5	450	9	AL799805	AL799805 AL799805
37	72.5	15.5	533	10	AW963120	AW963120 EST375193
38	72.5	15.5	544	9	AL898643	AL898643 AL898643
39	72.5	15.5	563	28	BZ424716	BZ424716 100018538
40	72.5	15.5	567	14	CF754312	CF754312 EST-77-2-
41	72.5	15.5	579	9	AL962576	AL962576 AL962576
42	72.5	15.5	601	29	CE253602	CE253602 tigr-gss-
43	72.5	15.5	640	9	AL846493	AL846493 AL846493
44	72.5	15.5	648	12	BM609423	BM609423 170006871
45	72.5	15.5	719	12	BG635313	BG635313 AT31857.5

ALIGNMENTS

RESULT 1  
LOCUS CF480372 584 bp mRNA linear EST 08-SEP-2003  
DEFINITION POL1\_65\_H07.b1\_A002 Pollen Sorghum bicolor cDNA clone  
ACCESSION POL1\_65\_H07\_A002 3', mRNA sequence.  
VERSION CF480372  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.





Db 152 GTGCCTTATCAAGTCCAGGAAGT---CCGGGTAGTAACTAATCTCCTCGTTCACTA 208

QY 27 LeuValAenThrGlySerSerSer----- 34

Db 209 GTATCAAAATTCGGCACCTCCTCACTTCCCTCGAAATGCCCTATAATTGAGTTTCGT 268

QY 35 -----ValAspLeuSerThrValThrValArgTyrTrpPheThr 47

Db 269 AAGGGGGAGCCTCCGAAGTTCTTGGTTATGAACATTTCTCCACTCGCAAAATGGGGTTCA 328

QY 48 ArgAspGly-----GlySerSerThr 54

Db 329 AGGGTTGGCTCAGGATCATTCACACCAAGTGGCTGGGCTCTAGCTAGTTCTGGAAC 388

QY 55 LeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74

Db 389 CTGACCCCAATGGTGGATTTCAAGGCTAGGTTCTGGTACTGTGCTCACTCCAAATGGTGG 448

QY 75 SerValAenProAlaThrProThrAlaAspThrTyrLeu 87

Db 449 GAA-----CCTCCTCCCGAGATAGTTACCTT 475

RESULT 3

LOCUS CB293085

DEFINITION CB293085 768 bp mRNA linear EST 28-FEB-2003

ACCESSION UCRCS01\_05cb12\_g1 Washington Navel orange cold acclimated flavedo & albedo cDNA library Citrus sinensis cDNA clone UCRCS01\_05cb12, mRNA sequence.

VERSION CB293085.1 GI:28618542

KEYWORDS Citrus sinensis

SOURCE Citrus sinensis

ORGANISM Citrus sinensis

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.

AUTHORS 1 (bases 1 to 768)

Close, T.J., Collura, K., Fenton, R.D., Feuerbacher, O., Kim, H.R., Kudrna, D., Wananaker, S., Wing, R. and Yu, Y.

TITLE Development of EST Resources and New Genetic Markers for California Citrus

JOURNAL Unpublished (2003)

CONTACT: Timothy Close

DEPARTMENT OF Botany & Plant Sciences, University of California

RIVERSIDE, CA, 92521-0124

TEL: 909/7873318

FAX: 909/7874437

EMAIL: timothy.close@ucr.edu

SEQ PRIMER: T3.

FEATURES

source

1..768

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Washington navel"

/db\_xref="taxon:2711"

/clone="UCRCS01\_05cb12"

/tissue\_type="rind containing flavedo and albedo"

/dev\_stage="Mature fruit"

/lab\_host="E. coli TJC121"

/clone\_lib="Washington Navel orange cold acclimated flavedo & albedo cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the field at University of California, Riverside Agricultural Operations since 1983. The scion was Washington Navel orange and the rootstock Carizzo Citrange. Tissue from mature fruit was harvested at mid-day in January 2002 during a cold spell, when pre-dawn temperatures were approximately -2 to 2 degree C. Approximately 2 cm median sections of the rind were excised in the field from several fruits, then wrapped in aluminum foil and frozen quickly in dry ice. Total RNA was extracted using a phenol extraction procedure described in J. Japanese Soc. Hort.

Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA library was made, and 1 million primary lambda cDNA clones were in vivo excised to give a population of phagescript SK(-) phagemids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside (Fenton). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3530 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wananaker) using the HarVest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Alignment Scores:

Pred. No.:	37.9	Length:	768
Score:	78.50	Matches:	20
Percent Similarity:	57.69%	Conservative:	10
Best Local Similarity:	38.46%	Mismatches:	15
Query Match:	16.74%	Indels:	7
DB:	14	Gaps:	2

US-09-917-376-5 (1-88) x CB293085 (1-768)

QY 3 GlyGlyValIysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLys 22

Db 612 GCGCGCATCAAC-----AAGAACACACACAGCAGCAGCAATAATAATAAAGTAAT 662

QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42

Db 663 AATAATAGCGACGTTGTACTAGTGGTAGCGTACGCTGGTGGTCAAGTGGCGTAGCG 719

QY 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54

Db 720 -----TTTTCGGAGCGGAGGAGGAGGAGCAGT 746

## RESULT 4

LOCUS BQ722991

DEFINITION BQ722991 906 bp mRNA linear EST 16-JUL-2002

ACCESSION AGENCOURT 8103733 Lupeki sympathetic\_trunk Homo sapiens cDNA clone

VERSION IMAGE:6191052 5', mRNA sequence.

KEYWORDS BQ722991.1 GI:21861888

SOURCE EST.

ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@email.nih.gov

Tissue Procurement: Dr. James R. Lupski

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM13590 row: n column: 13

High quality sequence stop: 512.

Location/Qualifiers

1..906

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

FEATURES

source

```

/clone="IMAGE:6191052"
/sex="male"
/tissue type="sympathetic trunk"
/dev stage="adult, 16 yr"
/lab host="DH10B"
/clone lib="Lupski sympathetic trunk"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dr priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGGCTCCG-3' and
5'-GACTAGTCTAGATCGGCGCGCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 73.1 Length: 906
Score: 77.00 Matches: 16
Percent Similarity: 49.15% Conservative: 13
Best Local Similarity: 27.12% Mismatches: 26
Query Match: 16.42% Indels: 4
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x BQ722991 (1-906)

QY 7 ValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGlyLeuGln 26
Db :||||| 16
548 CTGCCCTATACCAATAAATCTCCAGCGAGGCGCACCGTGAGTA-----ATGAAG 598
QY 27 LeuValAsnThrGlySerSerValAspSerThrValThrValArgTyrTrpPhe 46
Db :||||| 16
599 GCTGTAAGATGGGAATCAAGATATAGTGTGGAGCCTCTCCAGCCTACTGTGATC 658
QY 47 ThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db :||||| 16
659 TCCCGG--GGCGGATCGGCCAGCTCTACACCTGCATCTGGGCGAGTGTGGA 712

RESULT 5
AA043716
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 376)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
889549
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: eschwats@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 212.

FEATURES
source
1..376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3761053"
/db_xref="taxon:9606"
/clone="IMAGE:487339"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone lib="Soares pregnant uterus NbHPU"
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGAGAAATTCGGCGCCCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
and Eco RI sites of the modified p77T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.: 28.5 Length: 376
Score: 76.00 Matches: 20
Percent Similarity: 43.42% Conservative: 13
Best Local Similarity: 26.32% Mismatches: 23
Query Match: 16.20% Indels: 20
DB: 9 Gaps: 3

US-09-917-376-5 (1-88) x AA043716 (1-376)

QY 17 GlyAspAsnGlnLeuLysProGlyLeuValAsnThrGlySerSerValAsp 36
Db :||||| 16
64 GGGGTTCCAGCTGAGCCCTGT---GCTATAGANAGCGCGCTCTCCAGGGGTGAG 120
QY 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
Db :||||| 16
121 CTGGGAGGCTT---CTGCGGTCTGGAGTCCCGCGCATGGCGCCAGTTCCCGCAGAAC 177
QY 55 -----LeuValTyrAsn 58
Db :||||| 16
178 CCCTCCAGAGTCCCGCGGATGCACACAGAGAGGGGCTTGGAGTACTTCAGGC 217
QY 59 CysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74
Db :||||| 16
238 TGTGACGGGTGCGCTCGCGGTGTGGCAAGTGAAGTCTCTGTGGCG 285

RESULT 6
BE362891
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACDAG;
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 606)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and
Pratt, L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
On Jul 20, 2000 this sequence version replaced gi:9304448.
Contact: Cordonnier-Pratt MW
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

```



```

KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
1 (bases 1 to 657)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and
Pratt, L.H.
An EST database from Sorghum: dark-grown seedlings
TITLE EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTWix
High quality sequence start: 89
High quality sequence stop: 651
POLYA=NO.

FEATURES
source
1..657
Location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

ORIGIN
Alignment Scores: 62 Length: 657
Pred. No.: 76.00 Matches: 23
Score: 76.00
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.20% Indels: 10
DB: 10 Gaps: 4

US-09-917-376-5 (1-88) x BB362820 (1-657)

QY 5 ValLysValGlnTyrlsAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
::: :::::
DB 204 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGGAGCCTGTGTGCGCCGGC 263
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
::: :::::
DB 264 -----CGCGCGGAGAGAACCCCTGAAACTTCCAAACGATCGGTGCAAGGAAA 311

QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
::: :::::
DB 312 CCAATGTGGGATCTGCGAGGAGGAGGTTGAGCAACATGAATATATATATATGATGAA 371

QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
::: :::::
DB 372 GCTGTGTGTAAGTGTGGTACGTTGTTCTTTGAGACGGTGTCTGCTCTGTGTGTCG 431

QY 79 AlaThrProThrAlaAsp 84
::: :::::
DB 432 AGCTTCAGTCTCGTGAT 449

RESULT 9
CF487649 676 bp mRNA linear EST 08-SEP-2003
LOCUS POL1_45_C08_b1 A002 Pollen Sorghum bicolor cDNA clone
DEFINITION POL1_45_C08_A002 3', mRNA sequence.
ACCESSION CF487649

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CF487649.1 GI:34516518  
EST.  
Sorghum bicolor (sorghum)  
Sorghum bicolor

REFERENCE  
AUTHORS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 676)  
Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,  
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,  
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and  
Pratt, L.H.

TITLE  
JOURNAL  
COMMENT

EST database from Sorghum: pollen  
Unpublished (2003)  
Other ESTs: POL1\_45\_C08\_g1.A002  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.  
Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACGACC)  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1..676  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone="POL1\_45\_C08\_A002"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Pollen"  
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;  
Site 2: XhoI; The library was prepared from polyA+ RNA  
from pollen at the late vacuolated-vacuolated stage of  
development. Pollen was harvested from greenhouse-grown  
panicles of sorghum line Btx623. Panicles were removed  
from the flag leaf prior to emergence, when no detectable  
amylase is present in pollen of male-fertile lines. This  
stage represents pollen collected from anthers about 8-14  
days prior to anthesis. Double-stranded cDNA was cloned  
unidirectionally into different DraIII sites of the  
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,  
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA  
insert."

ORIGIN

Alignment Scores: 64.6 Length: 676  
Pred. No.: 76.00 Matches: 23  
Score: 76.00  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 14 Gaps: 4

US-09-917-376-5 (1-88) x CF487649 (1-676)

QY 5 ValLysValGlnTyrlsAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
::: ::::: |||||  
DB 214 ATAGCAATCAGCGCAAACTCAATGAATCCGGTCCGGAGGAGGAGCCTGTGTGCGCCGGC 273  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
::: ::::: |||||  
DB 274 -----CGCGCGGAGAGAACCCCTGAAACTTCCAAACGATCGGTGCAAGGAAA 321

Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 322 CCATGGTGGGGATCTGCAGGAGGAGGTTTGACCAACATGGAATATATATATGAA 381  
 Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 382 GCTTGTGGAAGTGTGGGACGTTGGTCTTCTTGGAGACGGTGTCTGTCTGTGGTCCG 441  
 Qy 79 AlaThrProThrAlaAsp 84  
 Db 442 AGCTTGCACTCTGGTAT 459

RESULT 10  
 BG739392  
 LOCUS EMI\_82\_D05.g1\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION BG739392  
 VERSION BG739392.1 GI:14089081  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 691)  
 REFERENCE Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.  
 An EST database from Sorghum: developing embryos  
 1 (bases 1 to 691)  
 TITLE Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude polyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 50  
 High quality sequence stop: 691  
 POLYA=No.

FEATURES  
 source  
 1. .691  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Embryo 1 (EM1)"  
 /note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."  
 Alignment Scores:  
 Pred. No.: 66.6 Length: 691  
 Score: 76.00 Matches: 23  
 Percent Similarity: 44.19% Conservative: 15  
 Best Local Similarity: 26.74% Mismatches: 38  
 Query Match: 16.20% Indels: 10  
 DB: 12 Gaps: 4

US-09-917-376-5 (1-88) x BG739392 (1-691)  
 Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 265 ATAGCAATCAGCCGAAATCAATCCGTCGCGGAGGAGCCTGTGTGGCCGCGGC 324  
 Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuThrValArg--- 43

Db 325 -----CGCGCGGGACAGAACCTTGAACCTTCARACGATCGGTGCAAGGAA 372  
 Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
 Db 373 CCATGGTGGGGATCTGCAGGAGGAGGTTTGACCAACATGGAATATATATGAA 432  
 Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
 Db 433 GCTTGTGGAAGTGTGGGACGTTGGTCTTCTTGGAGACGGTGTCTGTCTGTGGTCCG 492  
 Qy 79 AlaThrProThrAlaAsp 84  
 Db 493 AGCTTGCACTCTGGTAT 510

RESULT 11  
 BG739392  
 LOCUS BZ616323/c  
 DEFINITION 724 bp DNA linear GSS 16-JAN-2003  
 igs8e08.g1 WGS-ZmaysP (DH5a methyl filtered) Zea mays genomic clone  
 igs8e08.5, genomic survey sequence.

ACCESSION BZ616323  
 VERSION BZ616323.1 GI:27766878  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 724)  
 REFERENCE Rabinowicz,P.D., O'Shaughnessy,A.L., Ballia,V., Dedhia,N.,  
 Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,  
 Zutavern,T., McCombie,W.R. and Martienssen,R.A.  
 Genomic shotgun sequences from Zea mays (methyl-filtered)  
 Unpublished (2002)  
 TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: igs8 row: e column: 08  
 Seq primer: -21M13UnivRev  
 Class: shotgun  
 High quality sequence stop: 728.

FEATURES  
 source  
 1. .724  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ig58e08"  
 /lab\_host="DH5a"  
 /clone\_lib="WGS-ZmaysP (DH5a methyl filtered)"  
 /note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;  
 The vector was digested with XbaI and one nucleotide was  
 added by fill in in the recessive 3' end. The genomic DNA  
 was nebulized, end repaired, adaptor ligated and size  
 fractionated using sephadex. The resulting fragments were  
 between 0.8 and 3 kb and were cloned into the vector (.x/y  
 reads in M13mp19, .b/g reads in pUC19). The same ligation  
 was transformed into DH5a."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 71 Length: 724  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 16.20% Indels: 43  
 DB: 28 Gaps: 6  
 US-09-917-376-5 (1-88) x BZ616323 (1-724)

```

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
DB TCCGAGCGCTCAGATCCCATACAAAGGAGGAGGAAACAGCAAGCTCGCGCAGCTCGGCC 528
QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
DB GCGGCCAGCTGACGAGCTGCGGAGGAGCCAGCAGTATCTTATCAACGTTGCCGCC 468
QY 30 Thr-GlySerSerValAsp----- 36
DB ACGAGGAAGTTCTGCTCAACTCCCGGAGATCGTCGGAGCGTCCAGAAGCTGGTTTCGAG 408
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
DB GTCAATCCGATCGAGCCGCTAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 348
QY 53 rThrLeuValTyrAsnCysAspTIPAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
DB TCGTCCG-----ACGTGCTCATGGCGGCGCACGGGG----- 317
QY 73 eGlySerValAsnProAlaLaProThrAlaAspThr 85
DB -----CCGGCTCACCACACTTCTTCTCC 293

RESULT 12
BZ545807
LOCUS BZ545807 814 bp mRNA linear EST 11-JUL-2001
DEFINITION 602937951F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5101069 5',
ACCESSION BZ545807
VERSION BZ545807.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 814)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11243 row: f column: 14
High quality sequence stop: 792.
FEATURES             Location/Qualifiers
     source
     1..814
     /organism="Mus musculus"
     /mol_type="mRNA"
     /strain="FVB/N"
     /db_xref="taxon:10090"
     /clone="IMAGE:5101069"
     /lab_host="DH10B (T1 phage-resistant)"
     /clone_lib="NCI_CGAP_Li9"
     /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Alignment Scores:
Pred. No.:      83.6      Length:      814
Score:          76.00     Matches:     28
Percent Similarity: 44.83%  Conservative: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match:      16.20%  Indels:      19
DB:              12      Gaps:         4

US-09-917-376-5 (1-88) x BZ545807 (1-905)

```

```

US-09-917-376-5 (1-88) x BZ545807 (1-814)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB GAGGTGGAGCCGAGTATCGGAGCGGAAACACACAGATCGCAAGGACAATGAGCAG--- 89
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB -----AGTTCCCGAGTGGACCTCAGGACCTG--- 116
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59
DB :::::::::::::::::::: 117
CTCCGCTACTACAAACAGAGCGCGCGCTCTCACACGATCCACGCGGATGTACGGCTGT 176
QY 60 Asp-TpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAl 79
DB :::::::::::::::::::: 177
GACGTGGGTGGACGCGGCC-----TCCTCCGGGTACGACGAGGTGCGCA 224
QY 79 aThrProThrAlaAspThr 85
DB :::::::::: 225
TAGACGCGCTCGGATTACA 243

RESULT 13
BZ545807
LOCUS BZ545807 905 bp DNA linear GSS 16-DEC-2002
DEFINITION OGALR72TM ZM2.0.7.1.5 KB Zea mays genomic clone ZMMBMA0089L24,
ACCESSION BZ545807
VERSION BZ545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
WhiteLAW, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES             Location/Qualifiers
     source
     1..905
     /organism="Zea mays"
     /mol_type="genomic DNA"
     /strain="B73"
     /db_xref="taxon:4577"
     /clone="ZMMBMA0089L24"
     /clone_lib="ZM2.0.7.1.5 KB"
     /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.:      97      Length:      905
Score:          76.00     Matches:     34
Percent Similarity: 36.28%  Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match:      16.20%  Indels:      43
DB:              28      Gaps:         6

US-09-917-376-5 (1-88) x BZ545807 (1-905)

```





US-09-917-376-5 (1-88) x CG270748 (1-1001)

```

QY      2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db      418 TCCGAGGCGTCAGCATCCCATACAAGAGGCGCAACAGCAACGCTCGCGCAGCTGGCC 477
QY      17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db      478 GCGGCCCGAGCTGAGCAGCTGCGCGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 537
QY      30 Thr-GlySerSerSerValAsp----- 36
Db      538 ACGAGGAAGTTCGTCAACTCCCGGAGATCGTCGGAGCGGTCCAGAAGGCTGGGTTTCGAG 597
QY      37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db      598 GTCATCCCGATCGAGCGCGGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 657
QY      53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db      658 TCGTGCG-----ACGTGCTCATGGGCGCGCACGGGG----- 688
QY      73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db      689 -----CCGGCCTCACCACACTTCTCTTCC 712

```

Search completed: May 12, 2004, 11:39:02  
 Job time : 828.547 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 04:43:36 ; Search time 21.976 Seconds  
(without alignments)  
2222.227 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSNPATPTADTYLQ 88

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO\_pool/US09917376/runat\_11052004\_114536\_28265/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QMT=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 SCGN 1 1 116 @runat\_11052004\_114536\_28265 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCUTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	195.5	41.7	2029	US-09-136-574A-46	Sequence 46, Appl
2	194.5	41.5	6416	US-09-136-574A-2	Sequence 2, Appl
3	192.5	41.0	11707	US-09-136-574A-1	Sequence 1, Appl
4	182.5	38.9	2977	US-07-862-588B-1	Sequence 1, Appl
5	165	35.2	1438	US-09-339-159B-3	Sequence 3, Appl
6	165	35.2	1482	US-09-198-956-9	Sequence 9, Appl
7	165	35.2	1482	US-09-198-955A-11	Sequence 11, Appl
8	165	35.2	1482	US-09-694-531-11	Sequence 11, Appl
9	165	35.2	1482	US-09-670-141-9	Sequence 9, Appl
10	165	35.2	1482	US-10-072-152-11	Sequence 11, Appl
11	120.5	25.7	1624	US-07-862-588B-5	Sequence 5, Appl
12	120.5	25.7	1775	US-07-862-588B-5	Sequence 5, Appl

13	114	24.3	486	1	US-08-048-164A-1	Sequence 1, Appl
14	114	24.3	486	1	US-08-048-164A-3	Sequence 3, Appl
15	114	24.3	486	1	US-08-460-462-1	Sequence 1, Appl
16	114	24.3	486	1	US-08-460-462-3	Sequence 3, Appl
17	114	24.3	486	1	US-08-460-457-1	Sequence 1, Appl
18	114	24.3	486	1	US-08-460-457-3	Sequence 3, Appl
19	114	24.3	486	1	US-08-460-458-1	Sequence 1, Appl
20	114	24.3	486	1	US-08-460-458-3	Sequence 3, Appl
21	114	24.3	486	2	US-08-460-455-1	Sequence 3, Appl
22	114	24.3	486	2	US-08-460-455-3	Sequence 3, Appl
23	114	24.3	486	2	US-08-330-394A-1	Sequence 1, Appl
24	114	24.3	486	2	US-08-330-394A-3	Sequence 3, Appl
25	114	24.3	499	3	US-09-006-636-5	Sequence 5, Appl
26	114	24.3	499	3	US-09-006-632-5	Sequence 5, Appl
27	114	24.3	499	4	US-09-325-274-5	Sequence 5, Appl
28	113	24.1	1146	3	US-09-277-716-21	Sequence 21, Appl
29	113	24.1	1146	4	US-09-609-161B-21	Sequence 21, Appl
30	77	16.4	8107	4	US-09-335-586-3	Sequence 3, Appl
31	75.5	16.1	5714	4	US-09-820-312D-393	Sequence 393, App
32	72.5	15.5	20284	4	US-09-526-193A-21	Sequence 21, Appl
33	70	14.9	25165	4	US-09-453-702B-39	Sequence 39, Appl
34	67.5	14.4	2241	2	US-08-838-219B-20	Sequence 20, Appl
35	67.5	14.4	2241	3	US-09-233-336A-20	Sequence 20, Appl
36	67.5	14.4	2241	3	US-09-233-752A-20	Sequence 20, Appl
37	67.5	14.4	2241	3	US-09-402-036-20	Sequence 20, Appl
38	67.5	14.4	2241	4	US-09-904-226-20	Sequence 20, Appl
39	67.5	14.4	2370	2	US-08-838-219B-19	Sequence 19, Appl
40	67.5	14.4	2370	3	US-09-233-336A-19	Sequence 19, Appl
41	67.5	14.4	2370	3	US-09-233-752A-19	Sequence 19, Appl
42	67.5	14.4	2370	3	US-09-402-036-19	Sequence 19, Appl
43	67.5	14.4	2370	4	US-09-904-226-19	Sequence 19, Appl
44	67.5	14.4	118067	4	US-09-497-855A-32	Sequence 32, Appl
45	66	14.1	711	1	US-08-061-092A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136.574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.



```
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 3,15e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.35% Conservative: 16
Best Local Similarity: 43.53% Mismatches: 31
Query Match: 41.04% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-1 (1-11707)
Qy 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTTTGAAGGTACTATACAAAGCAATGAGACGAGCGACAAAGTCTCTAATAGGCG 4097
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGGTTTAAGATAGTCAATGGAGCGAGCAGCAGTGTGTAGCAGCGTTAAGATAAGA 4157
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 4158 TACTGGGTACACAGTGGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGGCAGCAGCATGTGACATTCAATTTGTGAGCTGAGCAGCGGAGTGAGTGGAGCG 4274
Qy 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.

; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
; US-07-862-588B-1

Alignment Scores:
Pred. No.: 1,05e-14 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.27% Conservative: 19
Best Local Similarity: 43.37% Mismatches: 25
Query Match: 38.91% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-1 (1-2977)
Qy 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGATGCACTACAGCGCGCGGATACAAATGCGAGCCGACACAGATCAAGCGCTCCTTC 203
Qy 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAAACAACCGGTACTTCGGCTGTGATTAAAGCAGCGCTCAAAATCCGCTACTAC 263
Qy 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGGATGGTTCTCGCGCGGTGAACGGCTGG---ATCGACTGGGCGCAGCTCGGC 320
Qy 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GGCAGCAACATTCAGATCTCGTTTGGC-----AACCATACTGCACGAATTCGATACG 374
Qy 86 TyrLeuGln 88
Db 375 TACGTGGAG 383
```

## RESULT 5

US-09-339-159B-3  
; Sequence 3, Application US/09339159B  
; Patent No. 6566114  
; GENERAL INFORMATION:  
; APPLICANT: Kauppinen, Markus  
; APPLICANT: Schulein, Martin  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Andersen, Lene  
; APPLICANT: Bjornvad, Mads  
; TITLE OF INVENTION: No. 6566114el Mannanases  
; FILE REFERENCE: 5440.204-US  
; CURRENT APPLICATION NUMBER: US/09/339,159B  
; CURRENT FILING DATE: 1999-06-24  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1438  
; TYPE: DNA  
; ORGANISM: Bacillus sp. 1633  
US-09-339-159B-3

Alignment Scores:  
Pred. No.: 8,98E-13 Length: 1438  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservative: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.18% Indels: 10  
DB: 4 Gaps: 2

US-09-917-376-5 (1-88) x US-09-339-159B-3 (1-1438)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20  
Db 940 GTATCAGGCAATTGGAAGTTGAACTTCAACAGCAATCCTTCAGATACTACTCA 999  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1000 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGAATGATTTGTCAAACTC 1059  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1060 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1116  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1117 CATGCTGCAATAATCGGCAGTACCGCAGCTACAACGGAATTACTTCAAAATGTAAGA 1176  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1177 ACATTGTAAATGAGTTCTCTCAACAAATAACGCAGACACCTACTCTTGAA 1227

## RESULT 6

US-09-198-956-9  
; Sequence 9, Application US/09198956  
; Patent No. 6165769  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels Erik K.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Schnorr, Kirk  
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus  
; FILE REFERENCE: 5377.200-US  
; CURRENT APPLICATION NUMBER: US/09/198,956  
; CURRENT FILING DATE: 1998-11-24  
; EARLIER APPLICATION NUMBER: 1344/97  
; EARLIER FILING DATE: 1997-11-24  
; EARLIER APPLICATION NUMBER: 60/067,240  
; EARLIER FILING DATE: 1997-12-02  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 9  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-198-956-9

Alignment Scores:  
Pred. No.: 9,37E-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservative: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.18% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-956-9 (1-1482)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGGCAATTGGAAGTTGAACTTCAACAGCAATCCTTCAGATACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGAATGATTTGTCAAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTACCGCAGCTACAACGGAATTACTTCAAAATGTAAGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTAAATGAGTTCTCTCAACAAATAACGCAGACACCTACTCTTGAA 1278

## RESULT 7

US-09-198-955A-11  
; Sequence 11, Application US/09198955A  
; Patent No. 6187580  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongbak, Lars  
; TITLE OF INVENTION: No. 6187580el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/09/198,955A  
; CURRENT FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-09-198-955A-11

```
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-955A-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTAATGTTGTCCTCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTACCTTGAA 1278

RESULT 8
US-09-694-531-11
; Sequence 11, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; CURRENT FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-694-531-11
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-694-531-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAGGTTACTAATACCGAAGCAGTGCATTAATGTTGTCCTCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTACCTTGAA 1278

RESULT 9
US-09-670-141-9
; Sequence 9, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-670-141-9
Alignment Scores:
Pred. No.: 9.37e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAAGGTTGAATCTTACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgala 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACAACGGAATTACTTCAAAATGTAAGAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTACCTTGAA 1278
```

1051	Db	ATCAATCCTCAGTTCAAGGTTACTAATACCGGAAGCAGTGCATTTGATTGTTCCAAACTC	1110
41	QY	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrIleuValTyrAsnCysAsp	60
1111	Db	ACATTGAGATATTATTATACAGTAGACGACAGAGAATCAGACCTTCTGG---TGTGAC	1167
61	QY	TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla	71
1168	Db	CATGCTGCAATAATCGGCAGTAAACGGCAGCTACACGGAATTACTCAATGTAAAAGGA	1227
72	QY	SerPheClySerValAsnProAlaThrProThrAlaAspThrTyrIleuGln	88
1228	Db	ACATTTGTAATAATAGATTCTCTCAACAAATAACGACACACCTACTCTTGA	1278

```

RESULT 10
US-10-072-152-11
; Sequence 11, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 537B.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

```

```

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

Qy      1 ValSerGlyGlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
          |||||
Db      991 GTATCAGCGCAATTGAAGGTTGAATCTCAACAGCAATCCTTCAGATACTACTAACTCA 1050
          |||||
Qy      21 IleLysProGlyLeuGlnLeuLysThrGlySerSerSerValAlaLeuLeuSerThrVal 40
          |||||
Db      1051 ATCAATCCTCAGTTCAGGGTTACTAATPACCGAGACGAGTCGAATGATTTGTCCAAACTC 1110
          |||||
Qy      41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

```

```

Db      1111 ACATTGAGATATTATTATACAGTAGACGGACAGAAAGATCAGACCTTCTCG--TGTGAC 1167
Qy      61  TPAAlaalaMet-----GlyCySGlyAsnIleArgAla 71
Db      1168 CATGCTGCATAATCGCAGTAACGGCAGCTACAACGGAATTACTTCAAAATGTAAGA 1227
Qy      72  SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db      1228 ACATTTGTAAAAATGAGTTCTCTCAACAAATAACGCAGACACCTTACCTTGAA 1278

RESULT 11
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch Iein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIM 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
; US-07-862-588B-6

Alignment Scores:
Pred No.: 1,17e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3

```

DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-6 (1-1624)

Qy 2 SerGlyGlyVallysValGlnTyrLysAenAsnAspSerAlaProGlyAspAenGinile 21  
Db 1323 ACGGGGAACCTTGTCCTCAATACAAAGTTGGCGCACACTAGCCGCCAGGATAACCAATG 1382  
Qy 22 LysProGlyLeuGlnLeuValAenThrGlySerSerValAspLeuSerThrValThr 41  
Db 1383 AGCCCTTCCTTTAAACATCAAGAACAACGGGTACAAACCCTGTTAACCTGAGCGGCCTCAAG 1442  
Qy 42 ValArgTyTrpPheThrArgAspGlySerSerThrLeuValTyrAenCysAspTyr 61  
Db 1443 CTTNNNNNNNNNNNNAAGAAC---GGACCTGCGGATATGAGCTCTCGATCGACTGG 1499  
Qy 62 AlaalaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAenProAlaThrPro 81  
Db 1500 GCCTTAATCGCGCAACGAATGTTCTGCTGGCATTC-----GCTAACTTTACCGGGAGT 1553  
Qy 82 ThrAlaAspThrTyr 86  
Db 1554 AATACGGATACTTAC 1568

RESULT 12

US-07-862-588B-5  
; Sequence 5, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linnae  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,588B  
; FILING DATE: 19920727  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 164/90  
; FILING DATE: 19-JAN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00013  
; FILING DATE: 18-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zelson, Steve T. / Lambiris, Elias J.  
; REGISTRATION NUMBER: 30,335 / 33,728  
; REFERENCE/DOCKET NUMBER: 3425.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1775 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus lautus



```
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 486 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: DNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..486
/ US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-1 (1-486)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 7 ACATCATCAATGTCAGTTGAATTTTACAACCTCTACAAATCAGCACAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 67 ACACCAATAATCAAAATTAACAACATCTGACAGTGATTTAAATTTAAATGACGTAAA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 127 GTTAGATATTATTACACAGTGATGTGTACACAGGACAAACTTTCTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 184 GCTGTGTCATTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

RESULT 14
US-08-048-164A-3/c
/ Sequence 3, Application US/08048164A
/ Patent No. 5496934
/ GENERAL INFORMATION:
/ APPLICANT: Shoseyov, Oded
/ APPLICANT: Shpiegl, Itai
/ APPLICANT: Goldstein, Marc A.
/ APPLICANT: Doi, Roy H.
/ TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: PENNIE & EDMONDS
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/048,164A
/ FILING DATE: 14-APR-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:

/ NAME: Mirock, S. Leslie
/ REGISTRATION NUMBER: 18,872
/ REFERENCE/DOCKET NUMBER: 7809-003
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 486 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: DNA
/ US-08-048-164A-3

Alignment Scores:
Pred. No.: 1.64e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-3 (1-486)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 480 ACATCATCAATGTCAGTTGAATTTTACAACCTCTACAAATCAGCACAACAACTCAATT 421
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 420 ACACCAATAATCAAAATTAACAACATCTGACAGTGATTTAAATTTAAATGACGTAAA 361
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 360 GTTAGATATTATTACACAGTGATGTGTACACAGGACAAACTTTCTGG---TGTGACCAT 304
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 303 GCTGTGTCATTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTC 244
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 243 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 196

RESULT 15
US-08-460-462-1
/ Sequence 1, Application US/08460462
/ Patent No. 5670623
/ GENERAL INFORMATION:
/ APPLICANT: Shoseyov, Oded
/ APPLICANT: Shpiegl, Itai
/ APPLICANT: Goldstein, Marc A.
/ APPLICANT: Doi, Roy H.
/ TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: PENNIE & EDMONDS
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/460,462
/ FILING DATE: concurrently herewith
/ CLASSIFICATION: 435
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-462-1

Alignment Scores:  
Pred. No.: 1.64e-06 Length: 486  
Score: 114.00 Matches: 29  
Percent Similarity: 52.58% Conservative: 22  
Best Local Similarity: 29.90% Mismatches: 34  
Query Match: 24.31% Indels: 12  
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-460-462-1 (1-486)

Qy	2	SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle	21
Db	7	ACATCATCAATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAACTCAATT	66
Qy	22	LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr	41
Db	67	ACACCAATATCAAAATTTACTACACATCTGCACGTGATTTAAATTTAAATGACGTAAA	126
Qy	42	ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp	61
Db	127	GTTAGATATTATTACACAAGTGATGGTACACAAGGACAACTTTCTGG--TGTGACCAT	183
Qy	62	AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe	73
Db	184	GCTGGTGCATTATTAGGAAATAGCTATGTTGATACACTAGCAAAAGTCACGCAAACTTC	243
Qy	74	-----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88
Db	244	GTTAAAGAAACAGCAGGCCCA---ACATCACTATGATACATATGTTGAA	291

Search completed: May 12, 2004, 15:15:30  
Job time : 26.976 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 12, 2004, 09:26:52 ; Search time 135.119 Seconds  
(without alignments)  
2950.286 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 459

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....TRASFGSVNATPTADTYLQ 88

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_11052004\_114537\_28336/app\_query.fasta\_1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376.@CGN 1 1 809 @runat\_11052004\_114537\_28336  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

RESULT 1  
US-09-917-378-2  
; Sequence 2, Appli  
; Publication No. US09917378  
; US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; CELLULOLYTICUS  
; FILE REFERENCE: 401977US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

ALIGNMENTS

1	459	100.0	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.4	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	99.4	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.5	2869	13	US-09-917-376-2	Sequence 2, Appli
5	462	98.5	2869	15	US-10-155-400-2	Sequence 2, Appli
6	257	54.8	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	190	40.5	1527	16	US-10-369-493-46838	Sequence 46838, A
9	165	35.2	1438	13	US-10-372-054-3	Sequence 3, Appli
10	165	35.2	1482	13	US-10-655-433-11	Sequence 11, Appl
11	165	35.2	1482	14	US-10-072-152-11	Sequence 11, Appl
12	157	33.5	1314	9	US-09-974-300-684	Sequence 684, App
13	155.5	33.2	4059	10	US-09-784-554B-1	Sequence 1, Appli
14	150.5	32.1	4056	10	US-09-784-554B-3	Sequence 3, Appli
15	130.5	27.8	4452	13	US-10-282-122A-15299	Sequence 15299, A
16	114	24.3	768	15	US-10-261-446-5	Sequence 5, Appli
17	113	24.1	1146	10	US-09-808-898-21	Sequence 21, Appl
18	78.5	16.7	3147	13	US-10-282-122A-13716	Sequence 13716, A
19	76	16.2	96597	12	US-10-052-482-226	Sequence 226, App
20	75.5	16.1	1041	16	US-10-260-238-742	Sequence 742, App
21	75.5	16.1	2091	13	US-10-211-462-222	Sequence 222, App
22	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
23	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appl
24	75.5	16.1	4041	13	US-10-312-312-21	Sequence 21, Appl
25	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
26	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App
27	75.5	16.1	4914	13	US-10-116-802-38	Sequence 38, Appl
28	75.5	16.1	5714	9	US-09-747-835A-19	Sequence 19, Appl
29	75.5	16.1	5714	9	US-09-747-835A-60	Sequence 60, Appl
30	75.5	16.1	5714	13	US-10-312-312-19	Sequence 19, Appl
31	75.5	16.1	5714	13	US-10-312-312-60	Sequence 60, Appl
32	75.5	16.1	5714	15	US-10-037-272-393	Sequence 393, App
33	75.5	16.1	5714	16	US-10-117-720-393	Sequence 393, App
34	75.5	16.1	5773	9	US-09-747-835A-18	Sequence 18, Appl
35	75.5	16.1	5773	13	US-10-312-312-18	Sequence 18, Appl
36	75	16.0	484	13	US-10-424-599-54043	Sequence 54043, A
37	73	15.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
38	72.5	15.5	402	9	US-09-764-860-85	Sequence 85, Appl
39	72.5	15.5	402	15	US-10-074-095-85	Sequence 85, Appl
40	72.5	15.5	402	16	US-10-212-872-85	Sequence 85, Appl
41	72.5	15.5	9181	10	US-09-984-827-10	Sequence 10, Appl
42	72.5	15.5	20284	13	US-10-617-334-21	Sequence 21, Appl
43	72.5	15.5	20284	16	US-10-452-510-21	Sequence 21, Appl
44	72	15.4	2402	13	US-10-425-114-9462	Sequence 9462, Ap
45	72	15.4	48667	9	US-09-822-268A-3	Sequence 3, Appli

## Alignment Scores:

Pred. No.: 2,21e-56 Length: 2289  
 Score: 469.00 Matches: 88  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-378-2 (1-2289)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 DB 1363 GGTGCGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATAACCGAG 1422  
 QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 1423 ATCAACCGGGTCTCCAGTTGGTGAATACGGGGTCTGCTCGGTGGATTTGTCGACGGTG 1482  
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1483 ACGGTGCGGTACTGGTTTACCCCGGATGGTGGTCTGCACACTGGTGTACAACTGTGAC 1542  
 QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1543 TGGCGCGGATGGGGTGTGGGAATATCCGGCCTCGTTCCGGTGGAAACCGCGGACG 1602  
 QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1603 CCGACGGCGGACACCTACCTGCAG 1626

## RESULT 2

US-09-917-384-2

; Sequence 2, Application US/09917384  
 ; Publication No. US20030096342A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917,384

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 3687

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of

; OTHER INFORMATION: Guxa

US-09-917-384-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
 Score: 466.00 Matches: 87  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.86% Mismatches: 0  
 Query Match: 99.36% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-384-2 (1-3687)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 DB 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATAACCGAG 1809  
 QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGCTCGGTGGATTTGTCGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1870 ACGGTGCGGTACTGGTTTACCCCGGATGGTGGTCTGCACACTGGTGTACAACTGTGAC 1929  
 QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1930 TGGCGCGGATGGGGTGTGGGAATATCCGGCCTCGTTCCGGTGGAAACCGCGGACG 1989  
 QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 3

US-09-917-383-2

; Sequence 2, Application US/09917383

; Publication No. US20030104522A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917,383

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 3687

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of

; OTHER INFORMATION: Guxa

US-09-917-383-2

## Alignment Scores:

Pred. No.: 1.13e-55 Length: 3687  
 Score: 466.00 Matches: 87  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.86% Mismatches: 0  
 Query Match: 99.36% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-383-2 (1-3687)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 DB 1750 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCGGTGATAACCGAG 1809

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGCTCGGTGGATTTGTCGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 1870 ACGGTGCGGTACTGGTTTACCCCGGATGGTGGTCTGCACACTGGTGTACAACTGTGAC 1929

QY 61 TtpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 1930 TGGCGCGGATGGGGTGTGGGAATATCCGGCCTCGTTCCGGTGGAAACCGCGGACG 1989

QY 81 ProThrAlaAspThrTyrLeuGln 88

DB 1990 CCGACGGCGGACACCTACCTGCAG 2013

## RESULT 4

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

```

; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 2,98e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 13 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTGGGTGGGTGGAGTGCAGTATAGAAATGATTTCGGCCGGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTCAGAGTGGTGAATATCCGGCTCGTTCGGTGAACCCGGCGAG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTACCCGGGATGGTGGCTCGTGCACACTGGTCAACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGGATCGGTGGTGGGAATATCCGGCTCGTTCGGTGAACCCGGCGAG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:

; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-5 (1-88) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTGGGTGGGTGGAGTGCAGTATAGAAATGATTTCGGCCGGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTCAGAGTGGTGAATATCCGGCTCGTTCGGTGAACCCGGCGAG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTACCCGGGATGGTGGCTCGTGCACACTGGTCAACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGGATCGGTGGTGGGAATATCCGGCTCGTTCGGTGAACCCGGCGAG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2223)
; OTHER INFORMATION:
US-10-156-761-550

Alignment Scores:
Pred. No.: 3,69e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 15 Gaps: 0

```

US-09-917-376-5 (1-88) x US-10-156-761-550 (1-2223)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 1771 TCGGCGGCGCTCAAGTCTCTTACAGAACACAGGACTCTTCGGCCACCGACGCGCATC 1830  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 1831 CGGCCAGGCGCTTCGGATCGTCAACCGGACGCGGCTCCCTCGACCTGTCCAAAGGTCAAG 1890  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 1891 GCCCGCTACTACTTACCCGGGACGCGCTCGCCACCGCTGCGGCGGACTAC 1950  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1951 GCGGCGGCTCGGCTGTTCACACGCTGTAAGTCTGTAACCTGACACCGCGCGGCGG 2010  
QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 2011 GGAGCGGACGCTACCTCGAA 2031

## RESULT 7

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 2,72e-21 Length: 9025608  
Score: 257.00 Matches: 49  
Percent Similarity: 71.26% Conservative: 13  
Best Local Similarity: 56.32% Mismatches: 25  
Query Match: 54.80% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-1 (1-9025608)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 706216 TCGGCGGCGCTCAAGTCTCTTACAGAACACAGGACTCTTCGGCCACCGACGCGCATC 706275  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 706276 CGGCCAGGCGCTTCGGATCGTCAACCGGACGCGGCTCCCTCGACCTGTCCAAAGGTCAAG 706335  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 706336 GCGCGCTACTACTTACCCGGGACGCGCTCGCCACCGCTGCGGCGGACTAC 706395

Db 706336 GCGCGCTACTACTTACCCGGGACGCGGCTCGCCACCGCTGCGGCGGACTAC 706395  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 706396 GCGGCGGCTCGGCTGTTCACAGCTCAGGCTGAGGTGCTACCCCTGACCAACCGCGTGC 706455  
QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 706456 GGAGCGGACGCTACCTCGAA 706476

## RESULT 8

US-10-369-493-46838  
; Sequence 46838, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 46838  
; LENGTH: 1527  
; TYPE: DNA  
; ORGANISM: Bacillus subtilis  
US-10-369-493-46838

## Alignment Scores:

Pred. No.: 8.02e-17 Length: 1527  
Score: 190.00 Matches: 35  
Percent Similarity: 65.12% Conservative: 21  
Best Local Similarity: 40.70% Mismatches: 26  
Query Match: 40.51% Indels: 4  
DB: 16 Gaps: 2

US-09-917-376-5 (1-88) x US-10-369-493-46838 (1-1527)

QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 1084 GGTATTTCTGTACAGTACAGCAGCGGATGGGATGATGAACAGCAACCAATCGTCGG 1143  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
Db 1144 CAGCTTCAATAAATAAATAACGGCAATACACGTTTGAATTTAAAGATGTCACTGCCCGT 1203  
QY 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62  
Db 1204 TACTGGTATAAAGCAAAACAAAGGCAAAAC-----TTTGACTGTGACTACGCG 1254  
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82  
Db 1255 CAGATTGGATGCGGCAATGTGACACACAAGTTTGTGCTGCAATAAACAAGCAAGGT 1314  
QY 83 AlaAspThrTyrLeuGln 88  
Db 1315 GCAGATACCTATCTCGAA 1332

## RESULT 9

US-10-372-054-3  
; Sequence 3, Application US/10372054  
; Publication No. US20030203466A1  
; GENERAL INFORMATION:  
; APPLICANT: Kauppinen, Markus  
; APPLICANT: Schuelein, Martin  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Andersen, Lene

```

; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-10-372-054-3

Alignment Scores:
Pred. No.: 2,72e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-372-054-3 (1-1438)
QY 1 ValSerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGCAATTTGAAGTTGAATTCACACAGCAATCCTTCAGATACTACTAACTCA 999
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAAGTTACTAATACCGGAAGCAGTGCAATGATTTGTCCAAACTC 1059
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGACGACGACGACGACGACGACGACGACGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTAGTACACGCAAGCAATTAATCAATGTAAAGGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACTCTTGAA 1227

RESULT 10
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217

```

```

; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-655-433-11

Alignment Scores:
Pred. No.: 2,84e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATTCACACAGCAATCCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTAATACCGGAAGCAGTGCAATGATTTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACGACGACGACGACGACGACGACGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAGTACACGCAAGCAATTAATCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACTCTTGAA 1278

RESULT 11
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217

```

```
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,846-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.73% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATTTCTACACAGCAATCTTCAGATCTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAAGGTTACTAATACCGAAGCAGTGAATTTGTTCCAAATC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATATACAGTAGCAGCAGAGAAGATCAGACCTCTCGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGACTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGAGCAGACACCTACCTTGAA 1278

RESULT 12
US-09-974-300-684
; Sequence 684, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-684

Alignment Scores:
Pred. No.: 3,346-12 Length: 1314
Score: 157.00 Matches: 32
Percent Similarity: 56.47% Conservative: 16
Best Local Similarity: 37.65% Mismatches: 33
Query Match: 33.48% Indels: 4
DB: 9 Gaps: 2

US-09-917-376-5 (1-88) x US-09-974-300-684 (1-1314)

; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 2,846-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.73% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 14 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATTTCTACACAGCAATCTTCAGATCTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCAGTTCAAGGTTACTAATACCGAAGCAGTGAATTTGTTCCAAATC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATATACAGTAGCAGCAGAGAAGATCAGACCTCTCGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGACTACAACGGAATTACTTCAAAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTCTCAACAAATAACGAGCAGACACCTACCTTGAA 1278

RESULT 13
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnoor, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.: 2,5e-11 Length: 4059
Score: 155.50 Matches: 27
Percent Similarity: 56.47% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 33.16% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-1 (1-4059)

QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGACTGCTGCTCCAATATCGCAGCAGATCTAAGGTGAACGATAATCACTCAATCCG 3657
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3658 CAATTCCAATTTGTAACAAGGCGACACCTCCATCCGATCCAGATGGAATTCGC 3717
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3718 TACTACTACAAATCGACGCGTACCGTGAGCAG--ACTTTCACTGCGACTATCGGACG 3774
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAGCTGAACGGTAAACTGTTAAATGGAGAGGCTGCAACCGGTGCC 3834
QY 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849

RESULT 14
US-09-784-554B-3
; Sequence 3, Application US/09784554B
```



```

; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schmorri, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLYCANASES
; FILE REFERENCE: 1001760-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 1.29e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.47% Conservativeness: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.09% Indels: 1
DB: 10 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-3 (1-4056)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGATGGTGGCTCCATATCGCACGGGATACAAATGTAACGACATCATTGAACCCG 3654
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCAAATTTTAAATAAGGTACAATCTCCGTAACCGATCAACGAGTTGAAATTCG 3714
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3715 TACTACTACGATCGACGGTGACCGTGAGCAG---ACATTCACATGCGACTATGCGGTG 3771
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3772 CTGAGCTCTCGAAGTGAATGGTAAGCTGGTTAAATGGATAAAGCTGCAACCGGTGCT 3831
Qy 84 AspThrTyrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 15
US-10-282-122A-15299
; Sequence 15299, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIORITY FILING DATE: 2000-05-26
; PRIORITY APPLICATION NUMBER: 60/230,335
; PRIORITY FILING DATE: 2000-09-06
; PRIORITY APPLICATION NUMBER: 60/230,347
; PRIORITY FILING DATE: 2000-09-09
; PRIORITY APPLICATION NUMBER: 60/242,578
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY APPLICATION NUMBER: 60/253,625
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY APPLICATION NUMBER: 60/257,931
; PRIORITY FILING DATE: 2000-12-22
; PRIORITY APPLICATION NUMBER: 60/267,636
; PRIORITY FILING DATE: 2001-02-09
; PRIORITY APPLICATION NUMBER: 60/269,308
; PRIORITY FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15299
; LENGTH: 4452
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15299

Alignment Scores:
Pred. No.: 1.04e-07 Length: 4452
Score: 130.50 Matches: 30
Percent Similarity: 55.79% Conservativeness: 23
Best Local Similarity: 31.58% Mismatches: 33
Query Match: 27.83% Indels: 9
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-282-122A-15299 (1-4452)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 100 AATTCGGGTGTTCAATATACAAATTTGCTGTATACAAATAGTACCAACCATGAATCTAAT 159
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 160 GCTCTTAATTTAAATCAAAATATATATCTGGAGCACCTTTAGATTATTAACAACTTAA 219
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 220 TTAAGATACATATTTTACAGCTGATGTTACTCAGGATGAAATTTTGG---TGCACCAT 276
Qy 62 AlaAlaMetGlyCysGly-----AsnIleArgAlaSerPhe 73
Db 277 GCTGTATGCTTAATGTTTATACTACCAACAACTTACAACTATGTAGTGGTACTTTT 336
Qy 74 GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 337 GTAGCTATGGATAATGCAACAGCTACTGCTGATCAATATCTTTGAG 381

Search completed: May 12, 2004, 18:28:26
Job time : 827.119 secs

```

DATE

5/14/07

APPLICATION NUMBER

09/917,376

DOC CODE

\_\_\_\_\_

DOC DATE

10/6/05

DELIVER THE ATTACHED FILE/DOCUMENT TO THE TC  
SCANNING CENTER

CONTRACTOR: THE ATTACHED FILE/DOCUMENT MUST BE  
INDEXED AND SCANNED INTO IFW WITHIN 8 WORK HOURS;  
UPLOADING OF THE SCANNED IMAGES SHOULD OCCUR NO  
LATER THAN 16 WORK HOURS  
FOLLOWING RECEIPT OF THIS REQUEST

AFTER SCANNING, ORIGINAL DOCUMENTS SHOULD BE BOXED IN  
ACCORDANCE WITH INSTRUCTIONS